ESD-TR-86-096

Technical Report
756

Adaptive Noise Reduction in Aircraft Communication Systems

J.J. Rodriguez

20 January 1987

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Prepared for the Department of the Air Force
under Electronic Systems Division Contract F19628-85-C-0002.

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract F19628-85-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Thomas J. Alpert

Thomas J. Alpert, Major, USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

ADAPTIVE NOISE REDUCTION IN AIRCRAFT COMMUNICATION SYSTEMS*

J.J. RODRIGUEZ

Digital Signal Processing Group, MIT Research Laboratory of Electronics and Group 24, Lincoln Laboratory

TECHNICAL REPORT 756

20 JANUARY 1987

* This report is based on the thesis of the same title submitted to the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology in May 1986 in partial fulfillment for the degree of Master of Science.

LEXINGTON MASSACHUSETTS

ABSTRACT

In many military environments, such as fighter jet cockpits, the increasing use of digital communication systems has created a need for robust vocoders and speech recognition systems. However, the high level of ambient noise in such environments makes vocoders less intelligible and makes reliable speech recognition more difficult. One method of enhancing the noise-corrupted speech is adaptive noise cancellation. In previous research, this method was tested in a simulated cockpit environment, yielding impressive results. However, in new simulations, reflecting more realistic conditions, adaptive noise cancellation has been less successful. Spectral analysis of the data shows that the spectral concentration of the ambient noise, along with the microphone characteristics, has a significant effect on the performance of adaptive noise cancellation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ADAPTIVE NOISE REDUCTION IN AIRCRAFT COMMUNICATION SYSTEMS

1. INTRODUCTION

With the advent of digital communication systems in fighter aircraft, there has been an increasing interest in developing vocoders and speech recognition systems for use in aircraft. However, the high levels of ambient noise in such environments make vocoders less intelligible and make reliable speech recognition more difficult. Therefore, attention has been directed toward the problem of enhancing the pilot's noise-corrupted speech. One method of improving the signal-to-noise ratio is adaptive noise cancellation (ANC).

ANC is a noise-reduction method that assumes no *a priori* knowledge of the noise or speech characteristics. Therefore, this technique has been considered for the problem of noise reduction in the cockpit environment, where combat conditions can lead to highly variant noise conditions. This method uses multiple inputs: a primary signal and one or more reference signals. The primary signal contains the noisy speech that needs enhancement. A second sensor provides the reference signal, which ideally contains only the ambient noise and no speech components. (In this research, only one reference input was used.) Adaptive filtering techniques are applied to these two signals in order to reduce the noise level in the primary output.

In previous research by Harrison,[9] ANC was tested on data collected from a simulation of a fighter jet cockpit environment. This resulted in a reduction of the noise by the impressivc amount of 11 dB. However, in a subsequent publication, Darlington *et al.*,[4] claimed that Harrison's success was a result of his use of only one noise source. In an actual cockpit environment, they claimed, the noise is diffusely distributed. In such a noise field, with the primary and reference sensors separated by a few centimeters, the coherence between the primary and reference signals becomes very small at frequencies above about 1 kHz. Therefore, ANC should perform poorly in an actual cockpit environment, in which there is a significant amount of noise above 1 kHz.

In order to assess the performance of ANC in a more realistic environment, new simulations were performed at the Wright-Patterson Air Force Base near Dayton, Ohio. One of the issues studied in these experiments was the effect of using multiple loudspeakers for generating the ambient noise field. When ANC was found to perform poorly on the data collected from these experiments, a second series of tests was conducted at MIT Lincoln Laboratory in Lexington, Massachusetts. This time, much attention was given to the primary microphone characteristics. Using only one loudspeaker, two types of primary microphones were tried: the standard-issue gradient microphone and an omnidirectional microphone. In analyzing the data from these experiments, the speech signals were not used; only the primary and reference noise signals were studied. When the gradient microphone was used, ANC only reduced the noise by about 2 dB.

When the omnidirectional microphone was used, ANC reduced the noise by about 9 dB. Therefore, the microphone characteristics seem to be an important factor. A thorough spectral analysis of all the data clearly shows why this is the case.

This report is organized into six sections. The second section includes an introduction to the theory of ANC. Section 3 summarizes previous research of ANC in aircraft communication systems. In the fourth section, the new simulations are described. The fifth section presents the experimental results, including a spectral analysis of the data. Finally, concluding remarks are given, along with suggestions for future research.

2. ADAPTIVE NOISE CANCELLATION

One method of enhancing speech corrupted with additive noise is to pass the signal through a linear, time-invariant filter. If the statistics of the speech and noise are stationary and known, then Wiener filtering theory can be used. However, when there is no *a priori* knowledge of the speech or noise, or when their statistics are nonstationary, adaptive filtering can often be an effective alternative.

Adaptive noise cancellation[3,5,10,19,20,21] is a noise-reduction method that uses multiple inputs: a primary signal and one or more reference signals. For speech applications, the primary signal is the noisy speech that we want to enhance. The reference signals are obtained from auxiliary sensors, located in the same noise field, but isolated from the primary sensor. For simplicity, we shall only consider the case of one reference signal. To enhance the noisy speech, the reference signal is adaptively filtered and then subtracted from the primary signal. The resulting output will hopefully contain the undegraded speech with less noise than the primary signal.

In this section, we begin with the theoretical development of ANC. This is followed by a description of the LMS algorithm, one method of implementation. In the final section, several limitations of ANC are discussed.

2.1 THEORETICAL DEVELOPMENT

Figure 2-1 shows the basic model of adaptive noise cancellation.[19] Here, s is the primary speech signal and n_r is the reference noise signal. In this model, the reference noise, n_r, passes

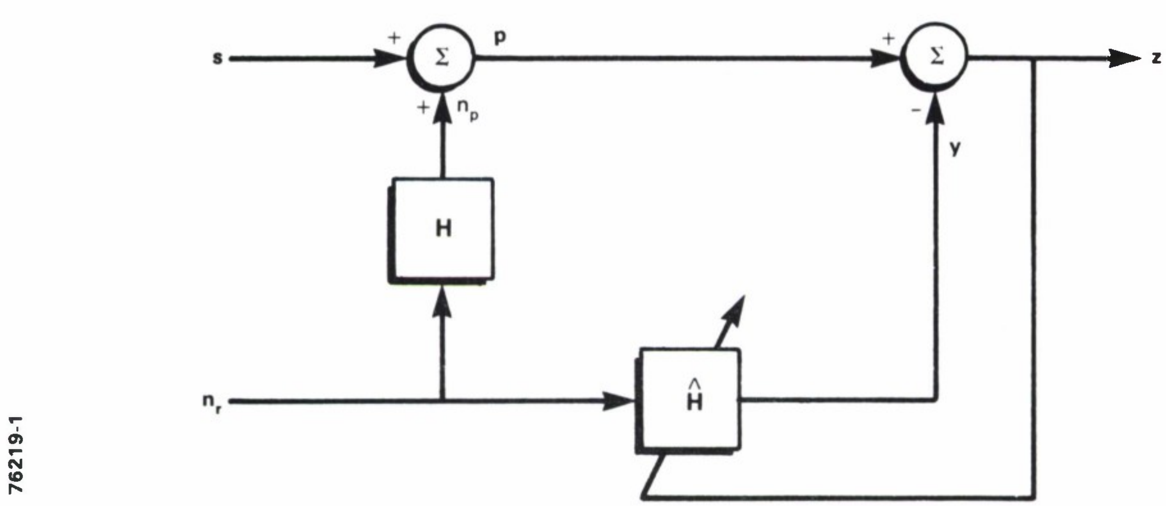


Figure 2-1. Basic ANC model.

through some transformation, H, to form the primary noise signal, n_p. In general, this transformation can be nonlinear and time-variant. However, the success of ANC depends upon the assumption that H is at least approximately linear. The primary signal (i.e., the noisy speech), p, is simply the sum of the primary speech and the primary noise signals. In order to enhance this primary signal, the reference noise is first processed by an adaptive filter, $\hat{H}$, resulting in an estimate, y, of the primary noise. Finally, this noise estimate is subtracted from the primary signal, yielding the enhanced output, z. In this model, we assume that s and n_r are uncorrelated random processes. If H were known *a priori*, then the trivial solution would be to let $\hat{H} = H$. Unfortunately, though, H is unknown in many cases of interest.

In order for the output, z, to be a minimum-mean-squared-error (MMSE) estimate of the desired signal, s, the adaptive filter must be varied so that the output noise power is minimized. For simplicity, assume that s and n_r are zero-mean, wide-sense stationary random processes. Because s is assumed to be uncorrelated with n_r, s is also uncorrelated with n_p and y. We can now compute the output noise power:

$$\begin{aligned} E[(z-s)^2] &= E[z^2] - 2E[sz] + E[s^2] \\ &= E[z^2] - 2E[s(s + n_p - y)] + E[s^2] \\ &= E[z^2] - E[s^2] - 2E[sn_p] + 2E[sy] \\ &= E[z^2] - E[s^2] \quad . \end{aligned} \tag{2.1}$$

The signal power, $E[s^2]$, is unaffected by the adaptive filter. Therefore, the output noise power is minimized by minimizing $E[z^2]$, the total output power. Equivalently, we can compute the mean squared error in terms of the error, $y - n_p$, in the noise estimate:

$$\begin{aligned} E[(z-s)^2] &= E[(s + n_p - y - s)^2] \\ &= E[(y - n_p)^2] \quad . \end{aligned} \tag{2.2}$$

This result shows that z is a MMSE estimate of s when y is a MMSE estimate of n_p, which agrees with intuition. Indeed, if it is possible to design $\hat{H}$ so that $y = n_p$ exactly, then the output will be noise-free, with $z = s$.

Thus, ANC can be viewed in several ways. In order for z to be a MMSE estimate of s, the adaptive filter coefficients must be varied in a particular manner. If they are chosen properly, then we will simultaneously realize the following equivalent results:

- z is a MMSE estimate of s;
- y is a MMSE estimate of n_p;
- The output noise power is minimized;
- The total output power is minimized.

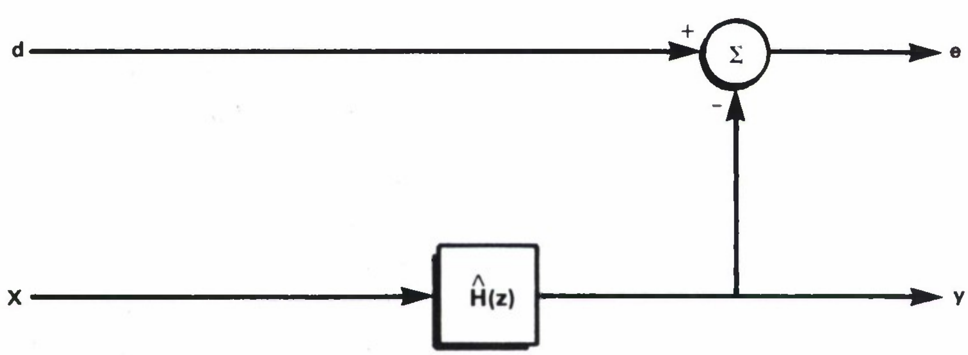
Minimization of the total output power is the basis for most ANC algorithms. The LMS algorithm was used to perform this minimization by continually updating the filter coefficients.

Suppose $\hat{H}$ is implemented as a linear, adaptive filter. Then perfect noise cancellation will only be possible if n_p and n_r are related by a linear filter. If n_p and n_r are correlated, but are not related by a linear filter, thereby violating the model in Figure 2-1, then the noise cancellation will only be partially effective. It is also instructive to consider the case in which the primary and reference noises are uncorrelated. In this case, the primary noise is uncorrelated with the filter output, y. Therefore, the output noise power becomes

$$\begin{aligned} E[(z-s)^2] &= E[(s+n_p-y-s)^2] \\ &= E[(n_p-y)^2] \\ &= E[n_p^2] - 2E[yn_p] + E[y^2] \\ &= E[n_p^2] + E[y^2] \quad . \end{aligned} \tag{2.3}$$

This is minimized by forcing all the filter coefficients to be zero, thereby shutting off the filter and causing $E[y^2] = 0$. The result is $z = p$. Therefore, when the input noises are uncorrelated, no noise cancellation occurs.

If s and n_r are stationary, and H is time-invariant, then we can use Wiener filtering theory to determine the optimal $\hat{H}(z)$, in terms of s and n_r. The classic form of a single-input, single-output Wiener filter is shown in Figure 2-2. Here, the input, x, passes through a linear, time-invariant filter, $\hat{H}(z)$. The filter is chosen so that the output, y, is an optimal estimate (in the least-squares sense) of the desired signal, d. By comparing Figure 2-1 and 2-2, we see that stationary ANC (with a linear $\hat{H}$) can be viewed as a Wiener problem, with $x = n_r$, $d = p$, and $e = z$. Of course, we must assume that the adaptive process has converged to the steady-state solution.



76219-2

Figure 2-2. Wiener filter.

It is well-known that the Wiener solution is given by the equation,

$$\hat{H}(z) = \frac{S_{n_r p}(z)}{S_{n_r n_r}(z)} \tag{2.4}$$

where

$$S_{n_r p}(z) = \sum_{m=-\infty}^{\infty} E[n_r(n)p(n+m)]z^{-m} \tag{2.5}$$

$$S_{n_r n_r}(z) = \sum_{m=-\infty}^{\infty} E[n_r(n)n_r(n+m)]z^{-m} \quad . \tag{2.6}$$

If n_r and n_p are exactly related by a linear filter, in accordance with the model in Figure 2-1, then the Wiener filter reduces to

$$\hat{H}(z) = \frac{S_{n_r n_r}(z)H(z)}{S_{n_r n_r}(z)} \tag{2.7}$$

$$= H(z)$$

as we would expect. In a general environment, where s and n_r may be nonstationary and $\hat{H}$ may be time-variant, $\hat{H}$ will tend to track H. However, the adaptive process takes time to converge. Therefore, successful tracking will occur only if the environment varies slowly.

2.2 LMS ALGORITHM

The adaptive filter, $\hat{H}$, is usually implemented as an FIR filter with variable coefficients. This is sometimes called an adaptive linear combiner or a tapped delay line. One of the more popular methods of updating the coefficients is the Widrow-Hoff LMS algorithm.[19] This algorithm is one of a class of algorithms that use the method of steepest descent to search for the optimal solution. The development of this algorithm begins with a general expression for the total output power as a function of the filter coefficients. In Section 2.1, we saw that optimal noise cancellation is achieved when the coefficients are adapted to minimize the total output power.

Let the impulse response of the adaptive filter be zero outside the interval, $0 \leqslant n < L$. To simplify the notation, define a time-variant vector of filter weights as

$$\underline{\hat{h}}_n = \begin{bmatrix} \hat{h}_0 \\ \hat{h}_1 \\ \cdots \\ \hat{h}_{(L-1)} \end{bmatrix}$$

where $\hat{h}_0, \hat{h}_1, \ldots,$ and $\hat{h}_{L-1}$ represent the filter coefficients and n is the time index. Similarly, we define a vector containing the latest L samples of the reference signal:

$$\underline{r}_n = \begin{bmatrix} r_n \\ r_{n-1} \\ \cdots \\ r_{n-(L-1)} \end{bmatrix}$$

Then the output of the adaptive filter at time n is the inner product or $\underline{r}_n$ and $\hat{\underline{h}}_n$:

$$y_n = \hat{\underline{h}}_n^T \underline{r}_n \quad . \tag{2.8}$$

The output of the system is given by

$$z_n = p_n - y_n = p_n - \hat{\underline{h}}_n^T \underline{r}_n \quad . \tag{2.9}$$

We must now adjust the filter weights so that the total output power is minimized. To do this, we first assume that the inputs are stationary and that the filter taps are fixed. Then the total output power, the "error function", is

$$\begin{aligned} E[z_n^2] &= E\left[(p_n - \hat{\underline{h}}_n^T \underline{r}_n)^2\right] \\ &= E\,[p_n^2] - 2E[p_n \underline{r}_n^T]\hat{\underline{h}}_n + \hat{\underline{h}}_n^T E[\underline{r}_n \underline{r}_n^T]\hat{\underline{h}}_n \quad . \end{aligned} \tag{2.10}$$

For simplicity, define $\underline{C}_n = E[p_n \underline{r}_n]$ and $\underline{R}_n = E[\underline{r}_n \underline{r}_n^T]$. Then the equation can be written as

$$E[z_n^2] = E[p_n^2] - 2\underline{C}_n^T \hat{\underline{h}}_n + \hat{\underline{h}}_n^T \underline{R}_n \hat{\underline{h}}_n \quad . \tag{2.11}$$

Therefore, the error function is a quadratic function of the weight vector, $\hat{\underline{h}}_n$. That is, this equation defines a hyperparaboloid in R^L. Because the output power is nonnegative, this surface must be concave upward. Consequently, there exists a unique global minimum, with no other local minima.

One way of searching for this minimum is the method of steepest descent. This is an iterative solution that continually updates the filter coefficients until the global minimum is reached. At each iteration, the filter taps are changed by an amount proportional to the negative gradient of the output power:

$$\hat{\underline{h}}_{n+1} = \hat{\underline{h}}_n - \mu \underline{\nabla}_n \quad . \tag{2.12}$$

Here, μ is an adaptation constant that controls stability and determines the rate of convergence. $\underline{\nabla}_n$ is the gradient of the error function at time n. This is obtained by differentiating Equation (2.11) with respect to the filter coefficients:

$$\underline{\nabla}_n = \left(\frac{\partial E[z_n^2]}{\partial \hat{h}_0}, \ldots, \frac{\partial E[z_n^2]}{\partial \hat{h}_{L-1}} \right)^T = -2\underline{C}_n + 2\underline{R}_n \hat{\underline{h}}_n \tag{2.13}$$

Instead of using this exact form of the gradient, the Widrow-Hoff LMS algorithm estimates the gradient by assuming that z_n^2 is a reasonable approximation of $E[z_n^2]$. Thus, we differentiate z_n^2 with respect to the filter coefficients:

$$\hat{\underline{\nabla}}_n = \left(\frac{\partial z_n^2}{\partial \hat{h}_0}, \ldots, \frac{\partial z_n^2}{\partial \hat{h}_{L-1}}\right)^T = 2z_n\left(\frac{\partial z_n}{\partial \hat{h}_0}, \ldots, \frac{\partial z_n}{\partial \hat{h}_{L-1}}\right)^T = -2z_n\underline{r}_n \tag{2.14}$$

The LMS algorithm is obtained by substituting this gradient estimate into Equation (2.12):

$$\hat{\underline{h}}_{n+1} = \hat{\underline{h}}_n + 2\mu z_n\underline{r}_n \quad . \tag{2.15}$$

This provides an easy way to iteratively compute an approximation of the optimal filter coefficients. For comparison, the ideal solution is easily obtained by setting the gradient of the error function to zero and solving for the filter weight vector. Using Equation (2.13), we obtain

$$\hat{\underline{h}}^* = \underline{R}_n^{-1}\underline{C}_n \quad . \tag{2.16}$$

This optimal filter weight vector is generally called the Wiener weight vector.

It can be shown that the gradient estimate used in the LMS algorithm is unbiased. However, the convergence of the filter weights is a much more complicated issue. If we make the assumption that the input vectors, $\underline{r}_n$, are stationary and uncorrelated over time, then the expected value of the vector of filter weights can be shown to converge to the Wiener weight vector:

$$\lim_{n\to\infty} \hat{\underline{h}}_n = \hat{\underline{h}}^* \quad . \tag{2.17}$$

However, this convergence is guaranteed only if

$$0 < \mu < \frac{1}{\lambda_{max}} \tag{2.18}$$

where λ_{max} is the largest eigenvalue of $\underline{R}$. ($\underline{R}_n$ is constant because $\underline{r}$ is stationary.) Rather than compute the eigenvalues, we can instead find an approximate upper bound for μ by considering the trace of $\underline{R}$. Because this is a positive semidefinite matrix, we have the following inequality:

$$\lambda_{max} \leqslant \sum_{i=1}^{L} \lambda_i = \mathrm{tr}[\underline{R}] \tag{2.19}$$

Because we are assuming stationarity of the reference signal, we have $\mathrm{tr}[\underline{R}] = LE[r^2]$, where $E[r^2]$ is simply the reference signal power. This leads to the following approximation for the bounds on μ:

$$0 < \mu < \frac{1}{\mathrm{tr}[\underline{R}]} = \frac{1}{LE[r^2]} \quad . \tag{2.20}$$

Although this analysis has assumed that the reference signal vectors are uncorrelated and stationary, the results seem to apply reasonably well in general practice. Unfortunately, no unconditional proof of convergence of the LMS algorithm is known at this time.

In addition to the bias, we must also examine the steady-state covariance of the filter weights. Widrow derived an approximate result,

$$\mathrm{cov}\,[\hat{\underline{h}}_n] \approx \mu\, \xi_{\min}\, \underline{I} \tag{2.21}$$

where ξ_{min} is the theoretical minimum value of $E[z_n^2]$ and $\underline{I}$ is the identity matrix. Therefore, a small adaptation constant results in less noise in the steady-state filter weights.

Unfortunately, a small adaptation constant also corresponds to slow convergence. The learning curve for the system output noise power can be approximated by a sum of exponentials. Widrow derived an estimate of the average time constant associated with this learning curve:

$$\tau \approx \frac{L}{4\mu \mathrm{tr}[\underline{R}]} \approx \frac{1}{4\mu E[r^2]} \tag{2.22}$$

Therefore, the time constant for convergence is inversely proportional to the adaptation constant. As a result, there is a trade-off between the steady-state covariance of the filter coefficients and the rate of convergence. However, if the environment is nonstationary, then a large adaptation constant must be chosen in order to provide adequate tracking of the environment. In this case, there is less freedom in choosing the adaptation constant.

2.3 LIMITATIONS

In this section, we address several factors that can degrade the performance of adaptive noise cancellation. These include several practical considerations, as well as two conditions that violate the basic model in Figure 2-1. One such violation is the presence of uncorrelated noises in the inputs. Another condition not accounted for in the model is the presence of speech components in the reference signal. As we shall see, these can seriously reduce the effectiveness of noise cancellation.

To begin with, there are several practical limitations associated with adaptive filtering. The LMS algorithm uses a causal, finite-extent, adaptive filter to estimate the transformation, H. However, H is not always modeled well by a causal filter. Therefore, it may be necessary to introduce a small delay into either the primary or the reference channel in order to achieve approximate causality. Another problem is limited frequency resolution. The frequency resolution is inversely proportional to the length of the impulse response of the adaptive filter. Consequently, there is a trade-off between the frequency resolution and the amount of computation.

In many applications, the convergence time of the adaptive filter is also a significant issue. Even in a stationary environment, the filter coefficients must undergo many iterations before they converge. This convergence time depends on several factors: the algorithm used to implement the adaptation, the length of the adaptive filter, and the shape of the reference noise spectrum.

In addition to these considerations, there are further limitations of ANC. In particular, the model of Figure 2-1 is often unrealistic. One condition that violates this model is the presence of uncorrelated noises in the primary and reference inputs.[19] To study this problem, let us first extend the original model to include uncorrelated noise sources, m_r and m_p. The modified model is shown in Figure 2-3. The primary signal now contains three components: $p = s + n_p + m_p$.

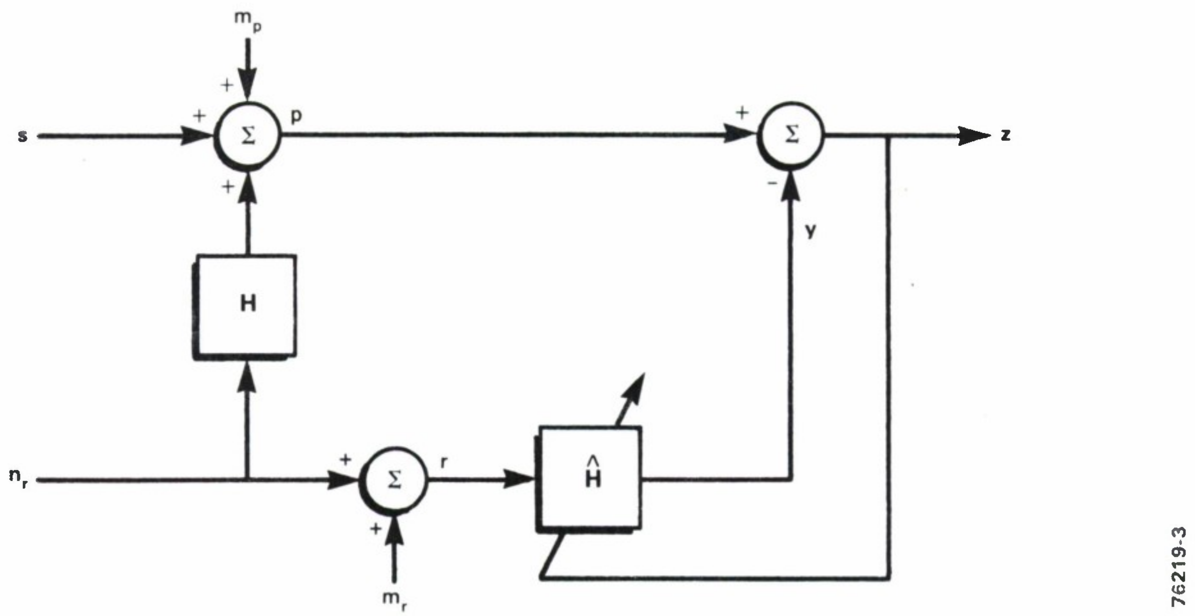


Figure 2-3. ANC model with uncorrelated noises.

Similarly, the reference signal, r, is now given by $r = n_r + m_r$. Next, define the signal-to-noise density ratio (SNDR), ρ, to be the ratio of the signal power spectral density to the noise power spectral density. This gives us a measure of the signal-to-noise ratio as a function of frequency. Assume that H is linear and time-invariant, and that all of the signals are stationary. For convenience, define the ratio of the uncorrelated noise spectrum and the correlated noise spectrum at the primary input to be

$$A(z) = \frac{S_{m_p m_p}(z)}{S_{n_p n_p}(z)} \quad . \tag{2.23}$$

Similarly, define the ratio of the uncorrelated noise spectrum and the correlated noise spectrum at the reference input to be

$$B(z) = \frac{S_{m_r m_r}(z)}{S_{n_r n_r}(z)} \quad . \tag{2.24}$$

Then it can be shown that the ratio of the output SNDR to the primary SNDR is

$$\frac{\rho_{out}(z)}{\rho_{pri}(z)} = \frac{[A(z) + 1]\,[B(z) + 1]}{A(z) + A(z)B(z) + B(z)} \quad . \tag{2.25}$$

This formula provides a measure of the performance of ANC in the presence of uncorrelated noises. It is important to note that if m_p and m_r are zero, then A(z) and B(z) are zero. From Equation (2.25), it is evident that, if this were the case, then there would be an infinite improvement in the SNDR. This is expected because, as was shown in the previous section, the Wiener solution could be used to give exact cancellation of the noise, resulting in a pure speech signal at the output. In general, the presence of uncorrelated noises diminishes the performance of ANC. From Equation (2.25), it is clear that the performance is maximized by minimizing A(z) and B(z), which measure the amount of uncorrelated noise present.

Another measure of the presence of uncorrelated noises is called the coherence.[16] The magnitude-squared coherence between two wide-sense-stationary random processes, x and y, is defined to be

$$\gamma_{xy}^2(z) = \frac{|S_{xy}(z)|^2}{S_{xx}(z)\, S_{yy}(z)} \tag{2.26}$$

where

$$S_{xy}(z) = \sum_{m=-\infty}^{\infty} E[x(n)y(n+m)]z^{-m} \tag{2.27}$$

$$S_{xx}(z) = \sum_{m=-\infty}^{\infty} E[x(n)x(n+m)]z^{-m} \tag{2.28}$$

$$S_{yy}(z) = \sum_{m=-\infty}^{\infty} E[y(n)y(n+m)]z^{-m} \tag{2.29}$$

(A more general definition would include a phase term. However, we shall only consider the magnitude.) For convenience, we shall refer to the magnitude-squared coherence simply as the coherence. It can be shown that the coherence, $\gamma_{xy}^2(z)$, represents the fraction of $S_{yy}(z)$ that is related to $S_{xx}(z)$ by a linear filter. If x and y are the input and output of a linear, time-invariant filter, then their coherence equals one. If y contains spectral components that are uncorrelated with x, then the coherence is less than one at the corresponding frequencies. If x and y are uncorrelated, then the coherence equals zero. Therefore, the coherence can be thought of as a correlation coefficient that is a function of frequency. Because ANC uses the reference noise to estimate the primary noise, a large coherence between the primary and reference noise signals is necessary if ANC is to be effective. In fact, an estimate of the amount of noise reduction can be given in terms of the coherence:

$$\frac{\rho_{out}(z)}{\rho_{pri}(z)} \leqslant \frac{1}{1 - \gamma^2_{n_r n_p}(z)} \tag{2.30}$$

where $\gamma^2_{n_r n_p}(z)$ is the coherence between the primary and reference noise signals. Therefore, it is clear that the performance increases as the coherence approaches one. It is also useful to define an attenuation function that measures the expected noise reduction in decibels:

$$\text{atten}(e^{j\omega}) = -10 \log_{10}[1-\gamma^2_{n_r n_p}(e^{j\omega})] \text{ dB} \tag{2.31}$$

In Section 5, the coherence and the attenuation function are used to analyze the performance of ANC with experimental data.

Another violation of the basic model is the leakage of speech components into the reference.[20] To determine the effect of this, let the model be extended to include a path from the primary input to the reference input. The transformation along this path will be denoted by H_2. For clarity, we shall now use H_1 to denote the transformation along the path from the reference to the primary. The resulting model is shown in Figure 2-4. For this discussion, the uncorrelated

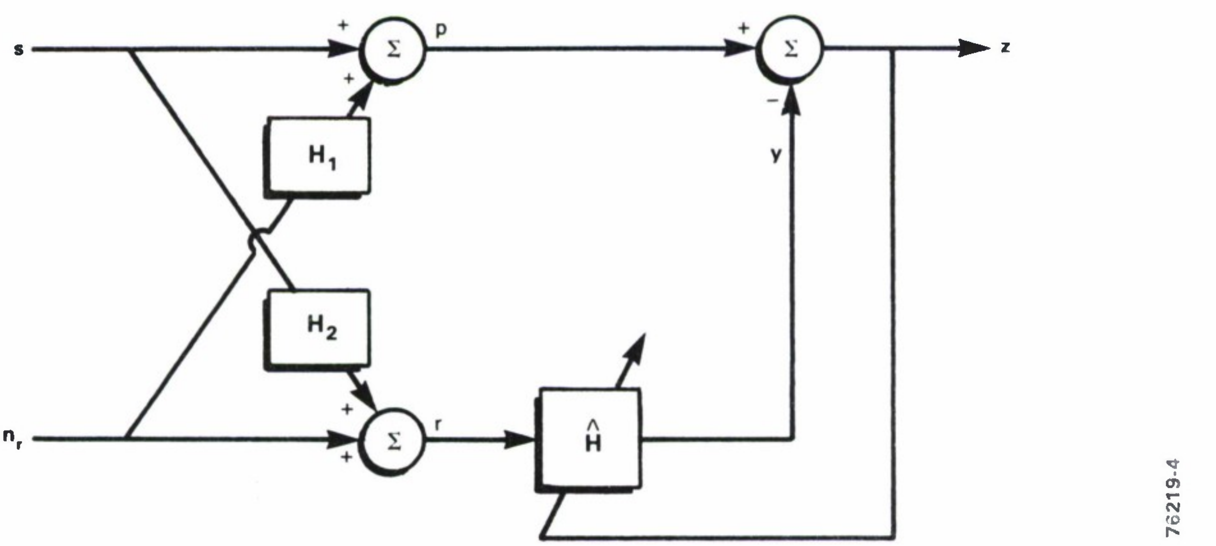


Figure 2-4. ANC model with speech leakage.

noises are omitted from the model. Of course, the environment could easily include uncorrelated noises in addition to speech leakage.

Widrow has shown that the resulting SNDR at the output is given by

$$\rho_{out}(z) = \frac{1}{\rho_{ref}(z)} \tag{2.32}$$

where $\rho_{ref}(z)$ is the SNDR at the reference input. When there is no leakage of speech into the reference input, this equation shows that the output SNDR will be infinite — i.e., the noise cancellation is exact. However, when the reference does contain speech components, the noise reduction will be only partially effective. Therefore, we see that the presence of speech in the reference does degrade the performance.

This degradation appears as distortion of the speech in the output, caused by cancellation of part of the speech signal. To evaluate the extent of this cancellation, define the signal distortion, D(z), to be the ratio of the spectrum of the speech component of y to the spectrum of the primary speech signal. Then it can be shown that

$$D(z) = \frac{\rho_{ref}(z)}{\rho_{pri}(z)} \tag{2.33}$$

Therefore, low signal distortion results from a low SNDR at the reference input and a high SNDR at the primary. This agrees with the intuitive behavior of the system.

As we have seen, there are many limitations of ANC. To begin with, there are several practical issues that need to be considered when implementing the technique. However, violations of the model are a much more serious issue. The presence of uncorrelated noises in the inputs places definite limits on the amount of noise reduction that can be expected. One useful measure of the amount of uncorrelated noise present is called the coherence. This measure is used later in the report to analyze the performance of noise cancellation. Another degradation occurs when speech leaks into the reference signal. This has the undesirable effect of canceling part of the speech in the output. Therefore, these limitations must be considered in any application of ANC.

3. ANC IN AIRCRAFT COMMUNICATION SYSTEMS

Adaptive noise cancellation has been successfully used in many applications. Only recently has it been considered for use in aircraft communication systems. The advent of digital communication systems in fighter aircraft has generated considerable interest in developing vocoders and speech recognition systems for use in aircraft. However, the high levels of ambient noise in such environments make vocoders less intelligible and make reliable speech recognition more difficult. Therefore, it has been proposed that ANC be used to enhance the pilot's noise-corrupted speech.

When ANC is applied to a fighter jet cockpit environment, numerous issues arise. For example, if the primary sensor is placed inside the pilot's oxygen face mask, where should the reference sensor be placed? Will the primary and reference noises be very correlated? What can be done if speech leaks into the reference input? Should the sensors be gradient microphones, or should they be omnidirectional microphones?

In the past, several researchers have used cockpit simulations to study the performance of ANC in aircraft. Harrison[9] was able to achieve significant noise reduction, but his simulation was very simplistic. Darlington *et al.*,[4] later showed that, in a diffuse noise field, the coherence between the primary and reference signals is very small above about 1 kHz. Therefore, they claimed that in an actual cockpit, ANC will only work well at very low frequencies. In this chapter, this past research is reviewed. But first, the fighter jet cockpit environment is described.

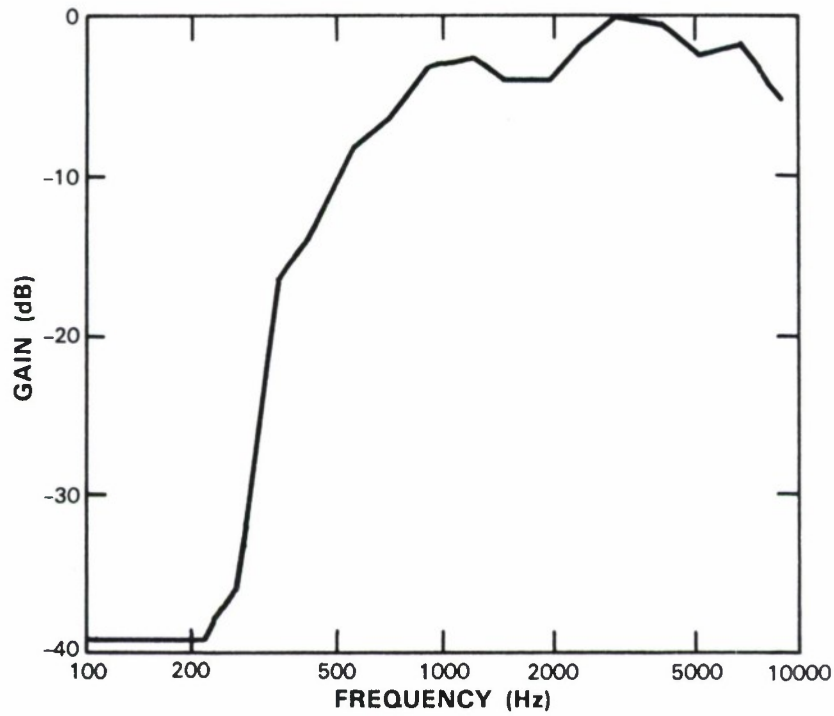
3.1 FIGHTER JET COCKPIT ENVIRONMENT

In a fighter jet cockpit, the pilot wears apparatus that provides him with a two-way communications link. The pilot wears an oxygen face mask, which is attached to his helmet. In turn, the oxygen face mask is connected to an oxygen supply via a flexible hose. Inside the face mask, a gradient, "noise canceling" microphone is mounted. This generates the primary signal in the model for adaptive noise cancellation. Also, a small acoustic speaker is mounted in each earpiece of the helmet. These speakers serve two purposes. First, they give the pilot the means to monitor radio communications. Also, they provide an audio feedback of the pilot's own speech. Without this feature, the pilot may find it difficult to hear himself speak, due to the high ambient noise level and the obstructing helmet.

In a nonradiating enclosure, such as an oxygen face mask, a gradient microphone offers performance superior to that of a pressure microphone. According to Morrow's studies of speech in nonradiating enclosures, the mask cavity tends to boost the low frequency energy and shift the formants (especially the first) upward in frequency.[12,13] Unlike a pressure microphone, a gradient microphone appears to counteract this bass boost, thereby removing the need for subsequent low-frequency equalization. Furthermore, the locations of the formants tend to be preserved more with a gradient microphone than with a pressure microphone. In addition to the change in formant frequencies, pressure microphones cause drastic changes in the relative amplitudes of the formants. Simple equalization of the pressure-microphone signal does not restore the locations of the formants; nor does it restore their relative amplitudes. Therefore, gradient microphones are preferred.

The relative superiority of gradient microphones in small, nonradiating enclosures is the result of two basic operating characteristics.[6,13] One highly desirable trait of gradient microphones is their directionality. The response to a near-field sound source is approximately proportional to the cosine of the angle of incidence (with respect to the axis of the microphone). Therefore, reflections from the sides of the face mask are de-emphasized. Another benefit of gradient microphones is their attenuation of far-field sound sources. This attenuation is quite significant at low frequencies, but becomes less pronounced with increasing frequency. In the cockpit application, the far-field sound is mostly ambient noise. Therefore, the gradient microphone becomes a noise-canceling microphone at low frequencies. If the power spectrum of the interfering noise decreases with frequency, then a significant amount of noise cancellation can result.

In the United States Air Force, the standard-issue oxygen face mask is equipped with an M-101 gradient microphone.[11] As shown in Figure 3-1, the far-field frequency response of this microphone peaks near 2.5 kHz. The response to frequencies below 1 kHz is much lower. In fact, the gain drops by about 36 dB as the frequency decreases from 1000 to 300 Hz.



76219-5

Figure 3-1. Frequency response of the M-101 gradient microphone.

3.2 PREVIOUS RESEARCH BY HARRISON

Several researchers have studied the application of adaptive noise cancellation to the fighter jet cockpit environment. The most successful results were obtained by Harrison[8,7,9] in a simulation of a fighter jet cockpit. The experiment was conducted in a partially soundproof room (about 10 ft by 10 ft) at MIT Lincoln Laboratory. In order to create the ambient noise field, digitally-created, white, gaussian noise was played through a loudspeaker, which was mounted on one of the walls. A subject, wearing a standard-issue oxygen face mask, was located near the loudspeaker. Two microphones were used to collect the data. One microphone, providing the primary signal, was placed inside the mask. A second microphone (Controlonics ME-9A electret-condenser type), providing the reference signal, was attached to the exterior of the mask, as close as possible to the primary microphone.

Two 2-channel recordings were then made. In the first 2-channel recording, speech was recorded, with the noise source turned off. In the second 2-channel recording, the ambient noise was recorded, with the subject holding his breath. Altogether, four signals were obtained: primary speech, reference speech, primary noise, and reference noise. The primary speech and primary noise signals were then digitally combined to form a composite primary signal. Similarly, the reference speech and reference noise signals were digitally combined to form a composite reference signal. The reason for recording the speech and noise separately was twofold. Because of the limited power of the loudspeaker, the primary noise signal was very small. Consequently, a reliable recording of a composite primary signal (speech plus noise) could not be obtained, due to the limited dynamic range of the recording equipment. Another reason for recording the speech and noise separately was that the signal-to-noise ratio (SNR) could be more easily manipulated.

Prior to Harrison's work, Boll and Pulsipher[1] used ANC to enhance noisy speech in an environment where no acoustic barrier was present. In order to keep the leakage of speech into the reference at a tolerable level, the primary and reference microphones had to be placed 12 ft apart. In the cockpit application, the oxygen face mask provides an acoustic barrier between the primary and reference. Because of this barrier, Harrison was able to place the microphones much closer to each other. Shortening the distance between the sensors is desirable for two reasons. It reduces the delay between the primary and reference signals, and it increases their coherence. In spite of the acoustic barrier, some of the speech still manages to leak into the reference. As was shown in Section 2.3, the presence of speech components in the reference signal results in less noise reduction, along with distortion of the processed speech signal, especially when the ambient noise level is low. To compensate for speech leakage, Harrison made a novel modification to the classic ANC method. Rather than update the adaptive filter after each input sample, he only updated the filter taps during speech inactivity. By incorporating a speech detection algorithm, he was able to freeze the filter taps during speech intervals and update them during silent intervals. Of course, the success of this technique depends on the stationarity of the noise. During speech intervals, the system must use a filter that was trained during the previous silent interval (up to 0.6 s earlier, according to Harrison). Furthermore, it is assumed that the silent intervals are long enough for the filter to converge. With the LMS algorithm, Harrison found that the convergence typically took about 120 ms.

In spite of these limitations, he was able to achieve very good results. Before processing the data, the primary signal was delayed by a small amount to force causality. By adding the appropriate delay, a shorter filter length could be used. Harrison found that a filter length of 50 taps was enough to increase the SNR by approximately 11 dB. Furthermore, the performance was approximately independent of the primary SNR. In addition to the LMS algorithm, Harrison tried the recursive least squares (RLS) algorithm for updating the filter coefficients. The RLS algorithm offers faster convergence at the expense of more computation. The two algorithms performed comparably.

3.3 PREVIOUS RESEARCH BY DARLINGTON ET AL.

Subsequent to Harrison's research, new findings were reported by Darlington *et al.*[4] They claimed that Harrison's simulation, which used only one loudspeaker, did not accurately represent an actual cockpit environment. Rather than modeling the noise field as a single noise source, they suggested modeling it as a diffuse noise field. In a diffuse noise field, the noise does not emanate from any one direction, but instead comes from independent sources from all directions.

The success of adaptive noise cancellation depends on the coherence between the primary and reference noise signals. However, it can be shown that, in a diffuse noise field, the coherence decreases as the spacing between the primary and reference sensors increases. This, of course, is the main reason for placing the reference microphone as close to the primary as possible. However, the coherence decreases not only with distance, but with frequency as well. It can be shown that, in a diffuse noise field, the coherence between the sound pressure levels at two points, r and p, is given by

$$\gamma^2_{rp}(e^{j\omega}) = \left[\frac{\sin(\omega d/c)}{\omega d/c}\right]^2 \tag{3.1}$$

where ω is the frequency, d is the distance between r and p, and c is the speed of sound.[16] (Here, no speech components are present.) The theoretical coherence for a typical spacing of 6 cm is shown in Figure 3-2. It is clear that there is little coherence above 1 kHz. In order to see the effect of this poor coherence on the performance of ANC, we compute the ideal attenuation function (from Equation 2.31):

$$\text{atten}(e^{j\omega}) = -10 \log_{10}[1-\gamma^2_{rp}(e^{j\omega})] \quad \text{dB} \tag{3.2}$$

Figure 3-3 shows a graph of the attenuation corresponding to the theoretical coherence in Figure 3-2.

76219-6

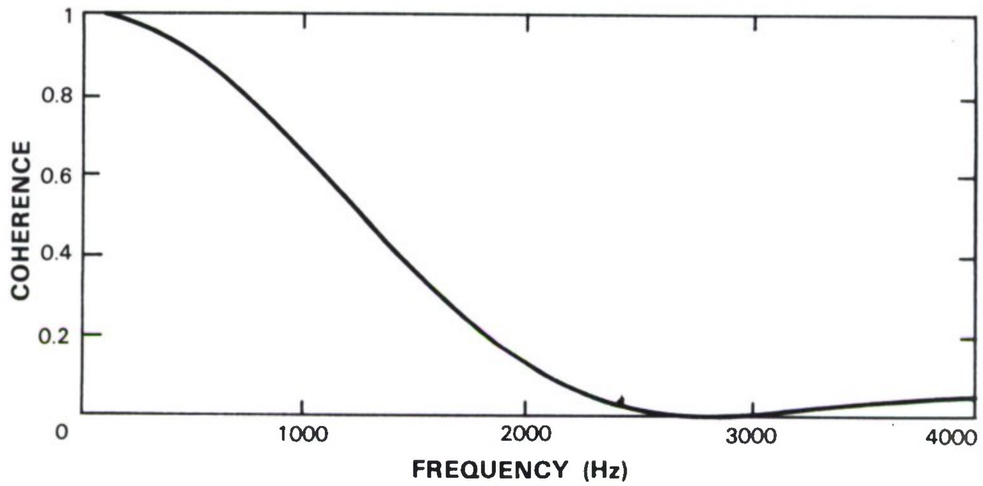


Figure 3-2. Theoretical coherence in a diffuse noise field.

76219-7

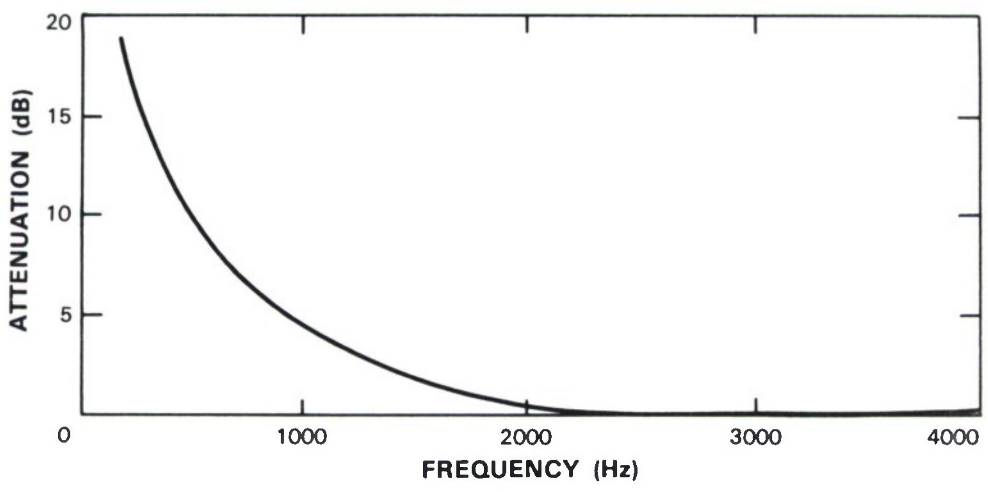


Figure 3-3. Theoretical attenuation in a diffuse noise field.

Darlington *et al.*, simulated a diffuse noise field in order to verify these theoretical results experimentally. Their simulation was similar to Harrison's, except that they used a British oxygen face mask and a more diffuse noise field. Using the data from this experiment, they estimated the coherence function. They found that the experimental coherence agreed quite well with the theoretical coherence. Therefore, they concluded that, in an actual cockpit, single-channel ANC will only be successful at very low frequencies.

The coherence function has thus been established as a valuable tool in the study of noise fields. However, these researchers did not measure the performance of ANC with their experimental data. While the relation between coherence and ANC has been shown theoretically, it has not been adequately studied experimentally. Also, the extent to which an actual cockpit environment agrees with the diffuse noise field remains an unanswered question.

Finally, one more point should be made. It has been shown that the coherence is high only for frequencies below about 1 kHz. If the primary signal is concentrated below 1 kHz, then ANC should work well. Therefore, the shape of the primary power spectrum is critical. In Section 5, experimental data supporting this is presented.

4. NEW SIMULATIONS

New simulations were conducted in order to extend Harrison's results to a more realistic environment. Previous simulations were limited in several respects. In the new simulations, much more attention was given to the microphone characteristics and the diffuseness of the noise. Two series of experiments were conducted. The first set of experiments took place at Wright-Patterson Air Force Base near Dayton, Ohio. The second set of experiments took place at MIT Lincoln Laboratory in Lexington, Massachusetts.

4.1 SHORTCOMINGS OF PREVIOUS SIMULATIONS

Harrison's simulation of a cockpit environment was kept very simple for practical reasons and for convenience. Some of the simplifications were not critical, whereas others probably affected his results.

To begin with, only one loudspeaker was used to generate the ambient noise. In an actual cockpit environment, the noise sources generally are not localized in any one area. Instead, the noise is scattered, emanating from all directions. Consequently, the noise field is modeled better by an array of loudspeakers, scattered in many directions, than by a single loudspeaker. Also, when only one loudspeaker is used, the direction of the loudspeaker, relative to the reference microphone, could affect the coherence between the primary and reference noise signals. By using multiple loudspeakers, placed in various directions, the effect of the orientation of the reference microphone is lessened.

In Harrison's experiment, the subject wore only the oxygen face mask, which was held in place by hand. The helmet and oxygen tank were disconnected. While worth mentioning, this modification probably had little effect on the results.

As previously discussed, the speech and noise were recorded separately and digitally combined later. Although this resulted in artificial signals, the signal-to-noise ratio could be manipulated easily without requiring additional recordings. The need for separate recordings was a practical one. Relative to the primary speech signal, the primary noise signal is usually quite small —i.e., the primary SNR is large (approximately 25 dB). Consequently, a reliable composite primary signal (speech plus noise) could not be obtained, due to the limited dynamic range of the recording equipment. It is unclear whether this artificial generation of the signals had any effect on Harrison's results.

In addition to these concerns, there are questions about the type of microphone used as the primary sensor. With Harrison's cooperation, a copy of his original data was made available for study. A comparison of this data with the data obtained from new experiments reveals evidence that Harrison may have used an omnidirectional microphone as the primary sensor. Unlike the standard-issue gradient microphone, the omnidirectional microphone does not offer any significant attenuation of far-field noise. Consequently, Harrison's primary noise signal contained a substantial amount of low-frequency energy. Estimation of the coherence between Harrison's primary and reference noises verified that only frequencies below about 1 kHz were highly correlated. Because of the high concentration of primary noise energy at low frequencies, ANC performed quite well. On the other hand, if a gradient, "noise-canceling" microphone had been used, then much of the low-frequency noise would have been canceled by the microphone, leaving little for adaptive noise cancellation to cancel. This issue is discussed in more detail in the next section.

4.2 SIMULATIONS AT WRIGHT-PATTERSON AIR FORCE BASE

In order to study the performance of adaptive noise cancellation in a more realistic cockpit environment, a series of new simulations were performed. The first set of experiments took place at Wright-Patterson Air Force Base (W.P.A.F.B.) near Dayton, Ohio. Recordings were made under a variety of conditions.

The experiments were conducted in a partially soundproof room (about 20 ft by 20 ft). Along two of the walls, numerous loudspeakers were mounted. A subject, wearing a helmet and an MBU-12/P oxygen face mask was seated near the center of the room. In front of him, an additional loudspeaker was placed. An M-101 gradient microphone, mounted inside the face mask, served as the primary sensor. The reference sensor, a Controlonics ME-9A omnidirectional electret-condenser microphone, was attached to the exterior of the mask, facing away from the mask. In addition, a small acoustic speaker was mounted in each earpiece of the helmet, providing an aural feedback of the subject's voice.

To generate the ambient noise, a Hewlett-Packard HP8057A precision noise generator was used. This noise was shaped with an equalizer before being played through loudspeakers. Two different types of noises were used. The first noise signal had an approximately flat power spectrum. The second noise signal was shaped to more closely resemble typical F-16 noise. These will be referred to as noise #1 and noise #2, respectively. However, it was later learned that the loudspeakers, which had a limited high-frequency power capability, were unable to accurately reproduce these noise signals. As a result, the two noises became very similar after passing through the loudspeakers. Both noise signals contained less high-frequency energy than was originally intended. Nevertheless, they proved to be adequate for the purposes of this research.

Several different configurations of the loudspeakers were investigated. Some of the experiments used only the single loudspeaker, located in front of the subject, while others used the many loudspeakers that covered two of the walls. When multiple loudspeakers were used, the noise was diffuse, emanating from many directions. When only one loudspeaker was used, the noise was concentrated in one direction, as it was in Harrison's experiment. However, this attempt to narrow the directionality of the noise source was hampered by the reverberation of the room. Because of the high noise intensity, reflections from the walls may have been significant. While no quantitative assessment of the reverberation of the room was made, it was apparent that the room was not designed to be anechoic.

Another item of interest was the oxygen tank. It is still unclear what effect its use may have on the performance of ANC. It was hypothesized that the passage of oxygen through the intake valve of the mask, together with an increased air pressure inside the mask, might accentuate the breath noise or introduce additional interference. This could have undesirable effects on the already-sensitive endpoint-detection algorithm used by Harrison. Therefore, experiments were performed both with and without the tank.

In most of these experiments, the speech and noise were recorded separately, using two 2-channel recordings, as Harrison had done in his experiment. However, a couple of experiments were also performed using composite recordings of speech and noise. That is, one 2-channel recording was made. One channel monitored the composite primary signal (speech plus noise), while the second channel monitored the composite reference signal (speech plus noise). This is more realistic than recording the speech and noise separately.

In order to study all of these variations, six experiments were performed. These corresponded to various combinations of the four variables: one loudspeaker/many loudspeakers, noise #1/noise #2, tank/ no tank, and separate/composite recordings. Specifically, the following experiments were performed:

Experiment 1: One loudspeaker, no tank, noise #1, separate recordings

Experiment 2: One loudspeaker, no tank, noise #2, separate recordings

Experiment 3: Many loudspeakers, no tank, noise #1, separate recordings

Experiment 4: One loudspeaker, no tank, noise #1, composite recordings

Experiment 5: One loudspeaker, tank, noise #1, separate recordings

Experiment 6: Many loudspeakers, tank, noise #2, composite recordings

When it was found that ANC performed poorly with even the simplest of these experiments, subsequent analysis of the data was limited to the first three experiments. Therefore, the effects of the oxygen tank and separate/composite recordings were not studied.

Several details concerning the recording procedure need to be mentioned. To generate a speech signal, the subject read a short paragraph. When the speech and noise were recorded separately, two 2-channel recordings were made. In the first recording, the primary and reference speech signals were recorded with the noise source turned off. In the second recording, the primary and reference noise signals were recorded while the subject was speaking and the noise source was turned on. In Experiments 1, 2, and 3, the noise recordings were made while the oxygen tank was still connected, for convenience. This should not matter because the subject was holding his breath, causing the intake valve of the mask to remain closed.

The noise levels should also be mentioned. When only one loudspeaker was used, the levels of noise #1 and noise #2 were both measured to be 106 dB SPL. When multiple loudspeakers were used, noise #1 was at 106 dB SPL and noise #2 was at 100 dB SPL.

4.3 SIMULATIONS AT MIT LINCOLN LABORATORY

When it was found that adaptive noise cancellation performed very poorly with the data from the simulations at Wright-Patterson Air Force Base, a second batch of experiments were performed. This time, the simulations took place at MIT Lincoln Laboratory in Lexington, Massachusetts, in the same room used by Harrison. The goal was to reproduce Harrison's results and explain why the W.P.A.F.B. data yielded such poor results.

In most of these experiments, the procedure was very similar to Harrison's. The subject wore only an MBU-5/P oxygen mask, which was held in place by hand. The helmet was not worn, and no oxygen tank was connected. Again, two microphones were used as sensors. The primary microphone was mounted inside the face mask. The reference sensor, a Controlonics ME-9A omnidirectional electret-condenser microphone, was attached to the exterior of the mask, facing away from the mask.

The primary issue being investigated was the effect of the environment on the coherence between the primary and reference noises. Therefore, only noise recordings were needed for the purposes of this research. Digitally-generated, white, Gaussian noise was used to create the ambient noise field. Again, the loudspeakers were limited in their high-frequency power capability. Consequently, the resulting noise field was concentrated at low frequencies.

In the first three of seven experiments, an M-101 gradient microphone was used as the primary sensor, and only one loudspeaker was used. The subject stood away from the loudspeaker by distances of approximately 1 ft, 4 ft, and 7 ft. Each time, the subject was oriented so that the reference microphone directly faced the loudspeaker.

In the next three experiments, an omnidirectional microphone (the same type as the reference microphone) was used as the primary sensor, and only one loudspeaker was used. It was known that Harrison had tried both the gradient and the omnidirectional microphones as primary sensors. However, there was some uncertainty about which type of microphone was used in the experiment that he reported. Therefore, three new experiments were performed using the omnidirectional microphone as the primary sensor. Again, the subject stood away from the loudspeaker by distances of approximately 1 ft, 4 ft, and 7 ft. Each time, the subject was oriented so that the reference microphone directly faced the loudspeaker.

In the last experiment, an M-101 gradient microphone was used as the primary sensor, and four loudspeakers were used. The loudspeakers were approximately located in each of the four corners of the room. This time, two independent noise sources (digitally-generated, white, Gaussian noise) were used. The first noise source was played through two of the loudspeakers at opposite corners of the room. The second noise source was played through the other two loudspeakers. The subject stood near the center of the room, where the noise level was 87 dB SPL.

Therefore, a total of seven experiments were performed at MIT Lincoln Laboratory:

Experiment 1: Gradient microphone, one loudspeaker, 1 ft

Experiment 2: Gradient microphone, one loudspeaker, 4 ft

Experiment 3: Gradient microphone, one loudspeaker, 7 ft

Experiment 4: Omnidirectional microphone, one loudspeaker, 1 ft

Experiment 5: Omnidirectional microphone, one loudspeaker, 4 ft

Experiment 6: Omnidirectional microphone, one loudspeaker, 7 ft

Experiment 7: Gradient microphone, four loudspeakers

5.0 EXPERIMENTAL RESULTS

As described in Section 4, many new experiments were performed. In the analysis of the experimental data, it is not sufficient to examine only the performance of adaptive noise cancellation. In order to gain a useful explanation of the variation in performance, it is also necessary to examine the power spectra of the signals involved.

Because of the decision to concentrate only on the noise components of the primary and reference signals, not all of the experimental data was analyzed. The noise signals from the first three experiments at Wright-Patterson Air Force Base and all seven of the experiments at MIT Lincoln Laboratory (L.L.) were chosen for the analysis. In addition, the noise signals from Harrison's experiment were also included. For convenience, these experiments will be referred to by the following names (for a more complete description of the experiments, see Sections 3.2, 4.2, and 4.3):

HAR: Harrison's experiment (L.L., one loudspeaker)

WP1: W.P.A.F.B., 1 loudspeaker, noise #1

WP2: W.P.A.F.B., 1 loudspeaker, noise #2

WP3: W.P.A.F.B., many loudspeakers, noise #1

LL1: L.L., gradient microphone, 1 loudspeaker, 1 ft

LL2: L.L., gradient microphone, 1 loudspeaker, 4 ft

LL3: L.L., gradient microphone, 1 loudspeaker, 7 ft

LL4: L.L., omnidirectional microphone, 1 loudspeaker, 1 ft

LL5: L.L., omnidirectional microphone, 1 loudspeaker, 4 ft

LL6: L.L., omnidirectional microphone, 1 loudspeaker, 7 ft

LL7: L.L., gradient microphone, 4 loudspeakers

5.1 PERFORMANCE OF ADAPTIVE NOISE CANCELLATION

In the first phase of processing the experimental data, the performance of adaptive noise cancellation was measured. For implementation, the LMS algorithm was chosen because of its computational efficiency and ease of use. Each of the eleven sets of data was processed according to the following procedure:

(1) The primary and reference noise signals were first passed through a 4 kHz anti-aliasing filter and digitized with a 16-bit analog-to-digital converter, using a sampling rate of 10 kHz.

(2) The signals were then passed through a 100 Hz high-pass digital filter, in order to eliminate 60 Hz hum and low-frequency drift.

(3) A 10-s segment was extracted from each signal and normalized to have zero mean and unit variance.

(4) The LMS algorithm was then applied, using a filter length of 100 samples. All of the filter coefficients were initialized to zero.

(5) Small amounts of delay were repeatedly added to either the primary channel or the reference channel until the amount of noise reduction was maximized.

(6) The adaptation constant was then incrementally varied until the maximum noise reduction was attained. In each case, $\mu = 0.003$ gave the best results.

In order to determine the amount of noise reduction, the output variance, σ_z^2, was estimated, using the last 9.5 s of the output signal. (Skipping the first 0.5 s allowed plenty of time for the adaptive filter to converge.) The noise attenuation in decibels was then given by $-10 \log_{10} \sigma_z^2$ (recall that the primary noise signal had unit variance). Table 5-1 shows the results of this performance analysis. Regarding these results, several observations can be made. First, the 15 dB attenuation achieved with Harrison's data is somewhat better than the 11 dB attenuation that was previously reported. This is easily explained. The 15 dB measurement was made using only the noise signals, whereas the 11 dB measurement was made using both the speech and noise signals. In all of the other experiments, one can see that ANC does, in fact, help reduce some of the noise. However, significant noise reduction was only obtained in the experiments in which an omnidirectional primary microphone was used (LL4, LL5, and LL6). In all of the experiments in which a gradient primary microphone was used, the noise attenuation was minimal. Therefore, there is a reason to suspect that Harrison's data was collected using an omnidirectional primary microphone. Finally, it should also be noted that the worst performance occurred in experiments WP3 and LL7, the only two experiments that used multiple loudspeakers. Based on these results, it appears that ANC only performs well when an omnidirectional microphone is used as the primary sensor. For an explanation of this peculiar result, we must turn to spectral estimation techniques.

5.2 METHOD OF SPECTRAL ANALYSIS

Spectral analysis of the experimental data offers further insight into the behavior of adaptive noise cancellation. In the analysis that was performed, several quantities were of interest:

(1) $S_{pp}(e^{j\omega})$, the power spectrum of the primary noise signal.

(2) $S_{rr}(e^{j\omega})$, the power spectrum of the reference noise signal.

(3) $S_{zz}(e^{j\omega})$, the power spectrum of the output noise signal.

(4) $\gamma_{rp}^2(e^{j\omega})$, the coherence between the primary and reference noise signals.

(5) $\hat{H}(e^{j\omega})$, the instantaneous transfer function of the adaptive filter.

TABLE 5-1

ANC Performance

Experiment	Noise Reduction (dB)
HAR	15.1
WP1	4.1
WP2	1.9
WP3	1.7
LL1	5.0
LL2	2.4
LL3	2.0
LL4	9.2
LL5	9.1
LL6	8.2
LL7	1.3

(Because the speech signals were not included in this analysis, $p = n_p$ and $r = n_r$.) The last item, $H(e^{j\omega})$, was easily obtained from the adaptive filter coefficients. However, for the other quantities, spectral estimation techniques were required.

Because the simulated noise signals were known to have smooth power spectra, it was decided that classical spectral estimation techniques would provide adequate resolution. In particular, the spectral estimates were obtained using Welch's method of averaging modified periodograms.[18] Consider, for example, the computation of $S_{pp}(e^{j\omega})$. In order to ensure stationarity, a segment of length 4096 was extracted from the primary noise signal. This 4096-point interval was then sectioned into 512-point segments, with a 256-point overlap between adjacent segments. In this manner, fifteen 512-point segments were obtained. We will refer to these as $p_k(n)$, where $0 \leqslant k \leqslant 14$ and $0 \leqslant n \leqslant 511$. The power spectrum estimate was then computed as follows:

$$S_{pp}(e^{j\omega}) = \frac{1}{15} \sum_{k=0}^{14} I_k(e^{j\omega}) \tag{5.1}$$

$$I_k(e^{j\omega}) = \frac{1}{512U} \left| \sum_{n=0}^{1023} p'_k(n)e^{-j\omega n} \right|^2 \tag{5.2}$$

$$p'_k(n) = \begin{cases} p_k(n)\, w(n) & , \quad 0 \leqslant n \leqslant 511 \\ 0 & , \quad 512 \leqslant n \leqslant 1023 \end{cases} \tag{5.3}$$

$$U = \frac{1}{512} \sum_{n=0}^{511} w^2(n) \tag{5.4}$$

where $w(n)$ was a Kaiser window. The computation of $S_{rr}(e^{j\omega})$ and $S_{zz}(e^{j\omega})$ followed the same procedure.

According to its definition, the coherence between the primary and reference noise signals is given by

$$\gamma^2_{rp}(e^{j\omega}) = \frac{\left|S_{rp}(e^{j\omega})\right|^2}{S_{rr}(e^{j\omega})S_{pp}(e^{j\omega})} \tag{5.5}$$

Here, $S_{rp}(e^{j\omega})$ is the cross power spectrum between the primary and reference noise signals. The Welch method for estimating the cross power spectrum is a straightforward extension of the method for estimating the auto-power spectra. The coherence estimate is obtained by substituting the Welch estimates of $S_{pp}(e^{j\omega})$, $S_{rr}(e^{j\omega})$, and $S_{rp}(e^{j\omega})$ into Equation (5.5) References 2 and 14.

Computation of $\hat{H}(e^{j\omega})$ was much simpler. The adaptive filter coefficients were observed at the middle of the time interval used for the power spectrum analysis. A 1024-point discrete Fourier transform of these filter coefficients was then computed, yielding the instantaneous transfer function of the adaptive filter.

In order to further reduce the variance of the power spectrum estimates, additional smoothing (low-pass filtering) of the estimates was performed. The low-pass filter that was used for this smoothing had a pass band edge at $\omega = \pi/8$. The computed transfer function of the adaptive filter was also smoothed by this amount. In addition, the coherence estimate was smoothed, using a low-pass filter with a pass band edge at $\omega = \pi/16$. The graphs of these smoothed estimates are much more visually pleasing than the unsmoothed estimates.

From the coherence estimate, the predicted attenuation function was computed. This is simply a more useful representation of the coherence function. It was computed from the smoothed coherence, according to the defining equation (from Equation 2.31):

$$\widehat{\text{atten}}(e^{j\omega}) = -10 \log_{10}[1 - \gamma_{rp}^2(e^{j\omega})] \text{ dB} \tag{5.6}$$

This provides a measure of the expected noise reduction, based upon the computed coherence function. Finally, the actual attenuation function was computed as follows:

$$\text{atten}(e^{j\omega}) = 10 \log_{10} \frac{S_{zz}(e^{j\omega})}{S_{pp}(e^{j\omega})} \tag{5.7}$$

By comparing the predicted attenuation with the actual attenuation, we can determine whether the coherence function is really a reliable predictor of ANC performance.

5.3 RESULTS OF SPECTRAL ANALYSIS

The spectral analysis was done for each of the eleven experiments. In Figure 5-1 to 5-11, the results are collected and displayed in graphical form. For each experiment, six graphs are shown: power spectra of the primary, reference, and output signals; the transfer function of the adaptive filter; the predicted attenuation function (derived from the coherence estimate); and the actual attenuation function. In each of the graphs, the abscissa represents the frequency, given in Hertz, and the ordinate value is plotted using a logarithmic scale (in decibels). Careful study of these graphs helps explain why adaptive noise cancellation performs well in some cases, but poorly in others.

Let us begin by studying the spectra of Harrison's data, shown in Figure 5-1. First, we see that the reference spectrum is not white; there is about a 10 dB/octave downward tilt. As mentioned before, this is probably due to the limited performance of the loudspeaker used in the simulation. The primary spectrum is even more concentrated at low frequencies, due to the high-frequency attenuation by the face mask. In fact, the band of frequencies below 1 kHz comprises 99.8 percent of the signal's energy. The low-pass characteristic of the face mask is also observed in the

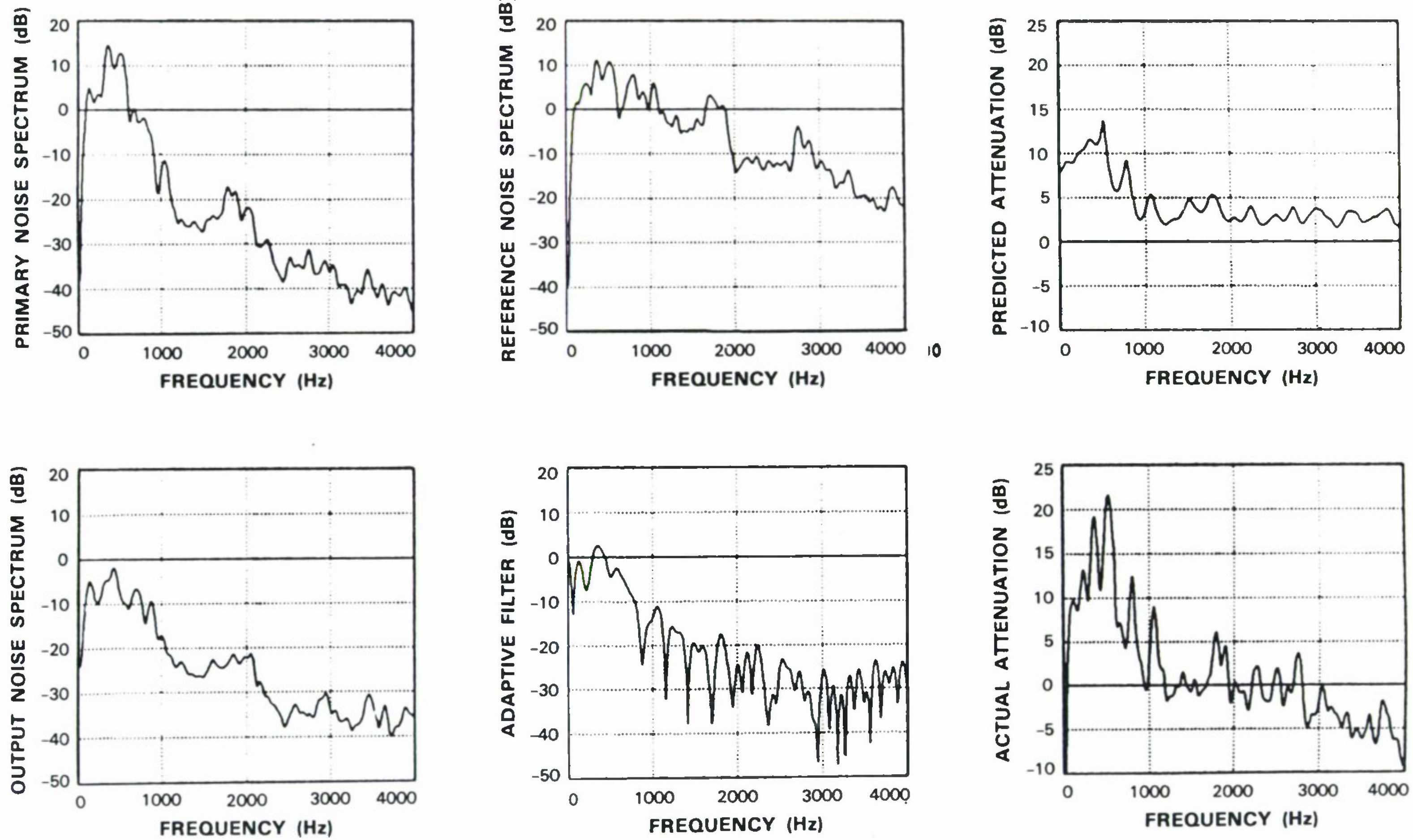


Figure 5-1. Spectral analysis: HAR.

76219-8

76219-9

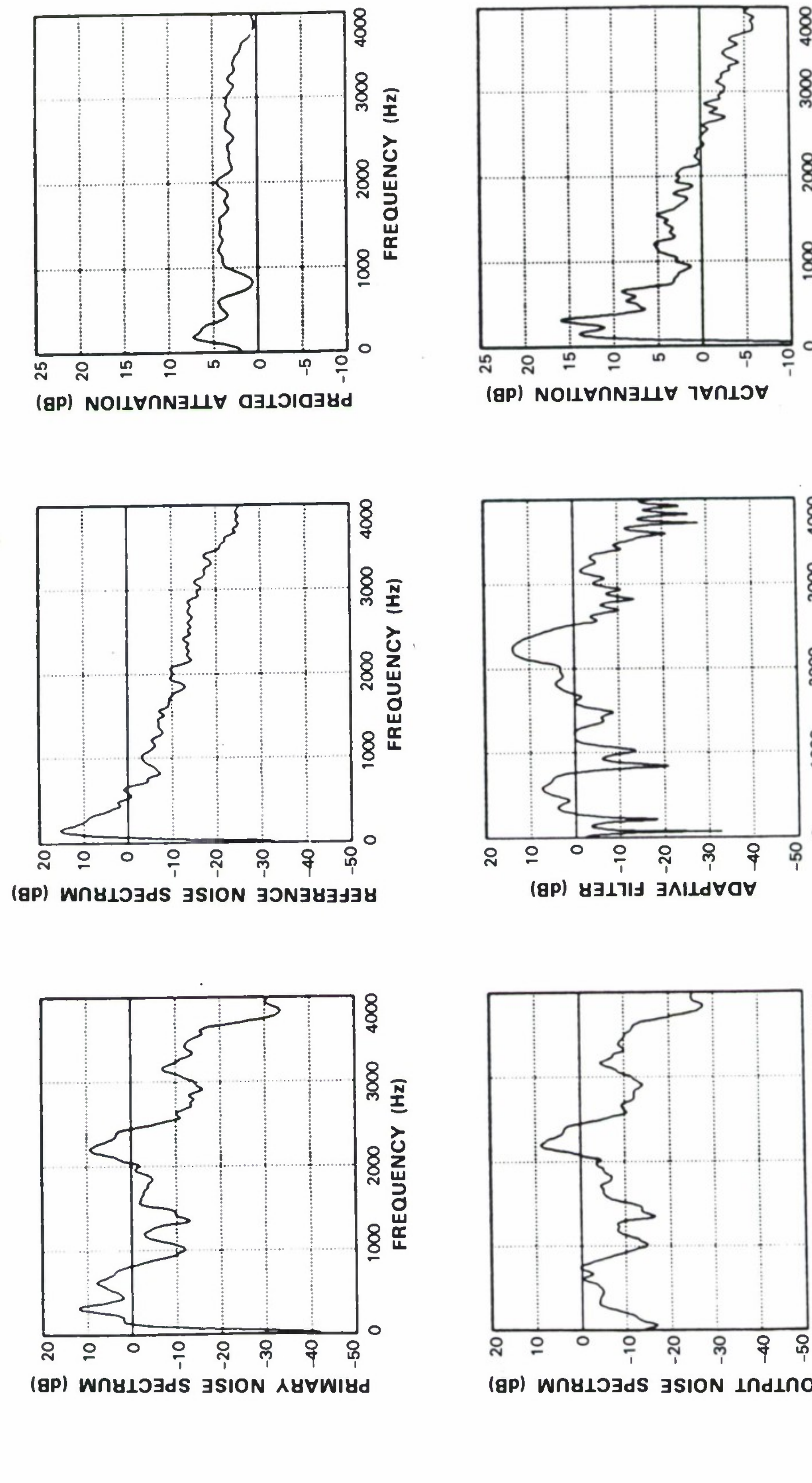


Figure 5-2. Spectral analysis: WP1.

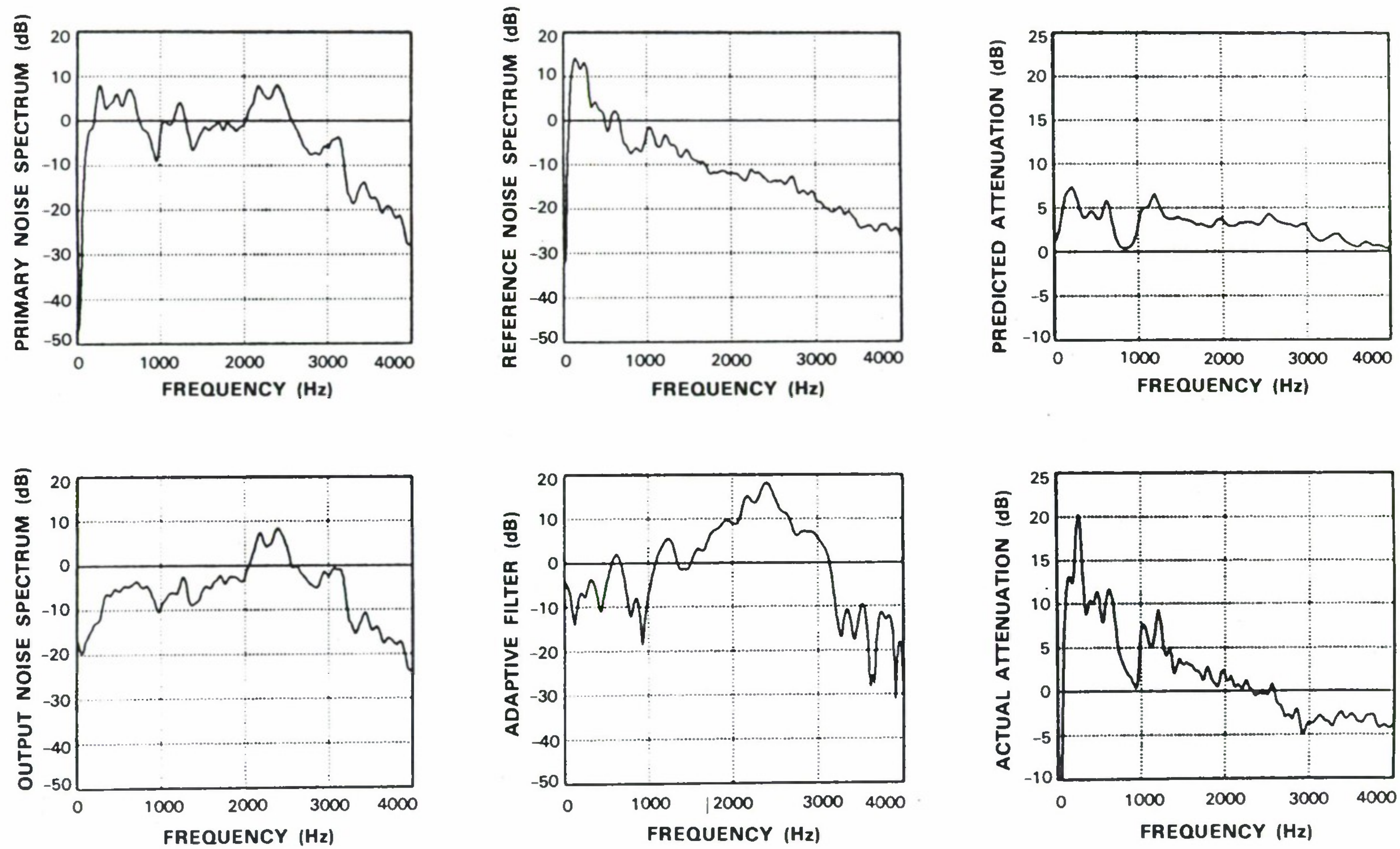


Figure 5-3. Spectral analysis: WP2.

76219-10

76219-11

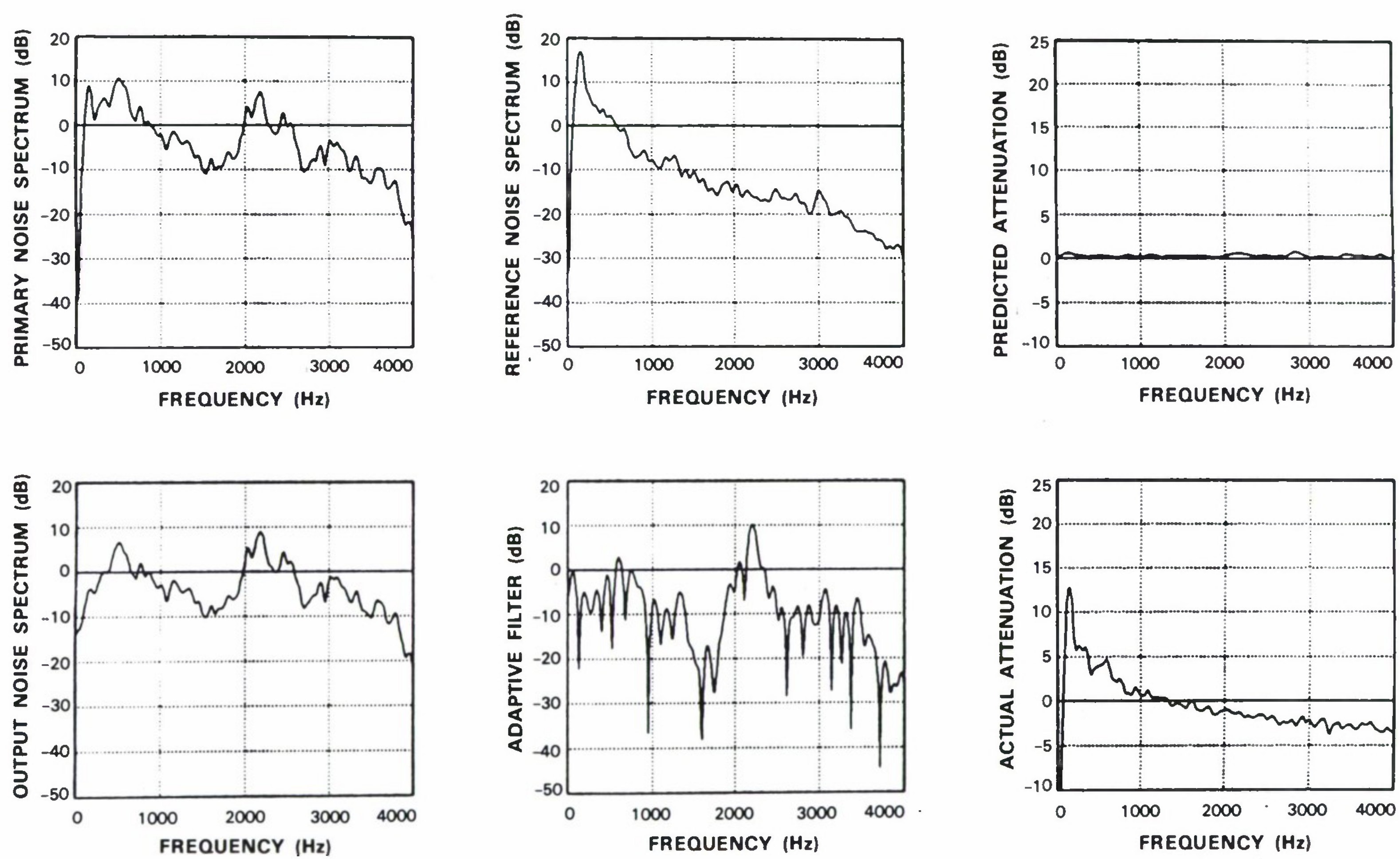


Figure 5-4. Spectral analysis: WP3.

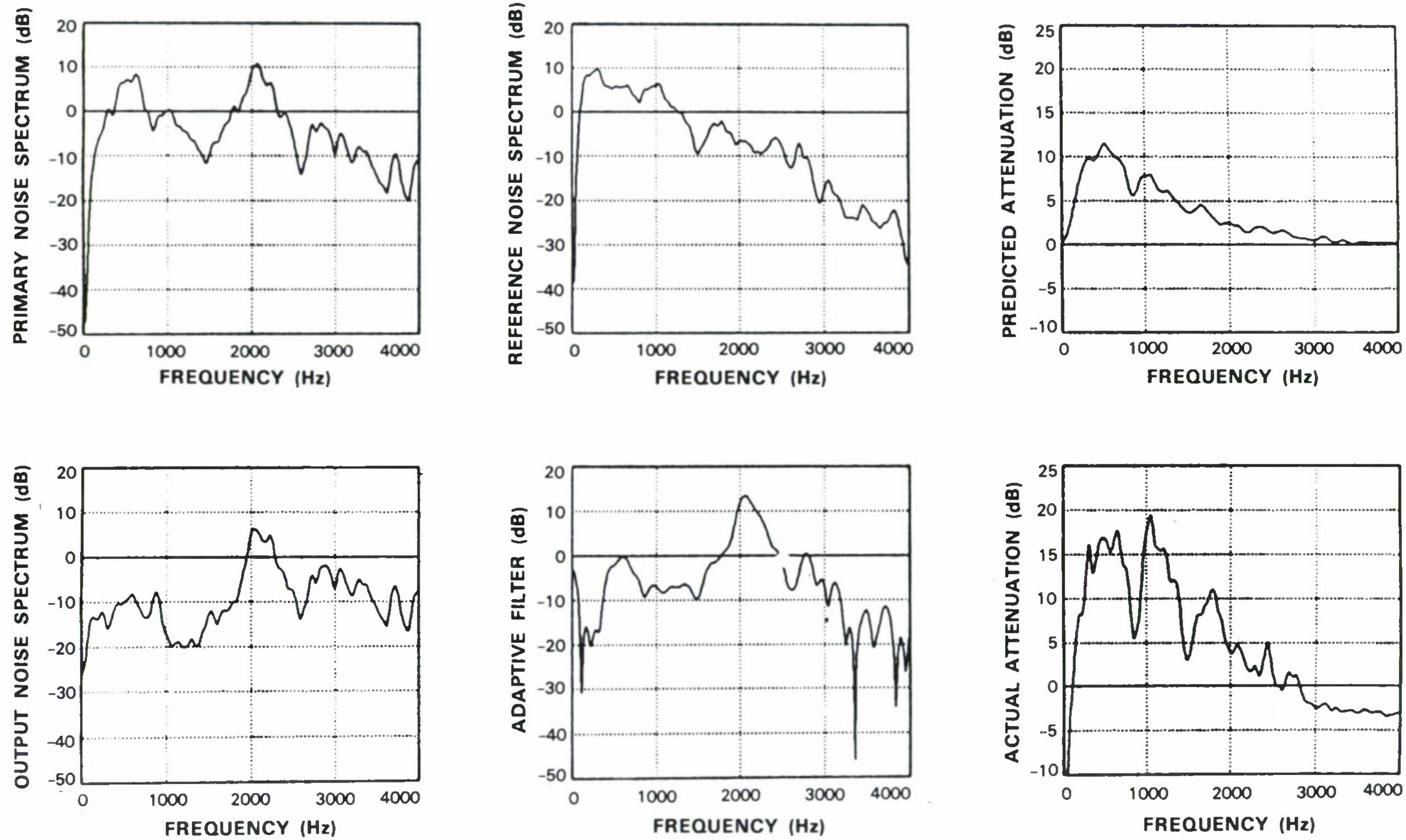


Figure 5-5. Spectral analysis: LL1.

76219-12

76219-13

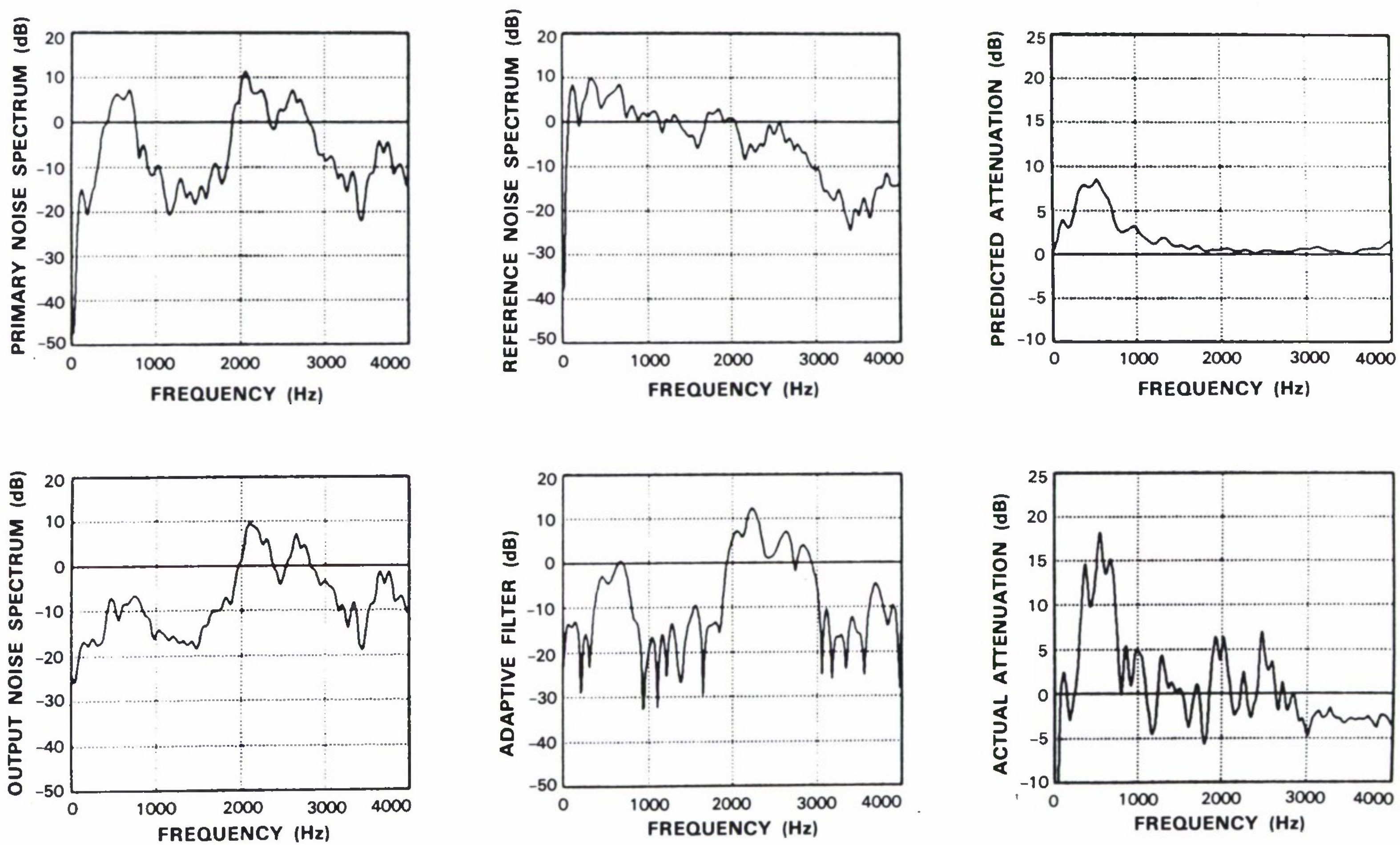


Figure 5-6. Spectral analysis: LL2.

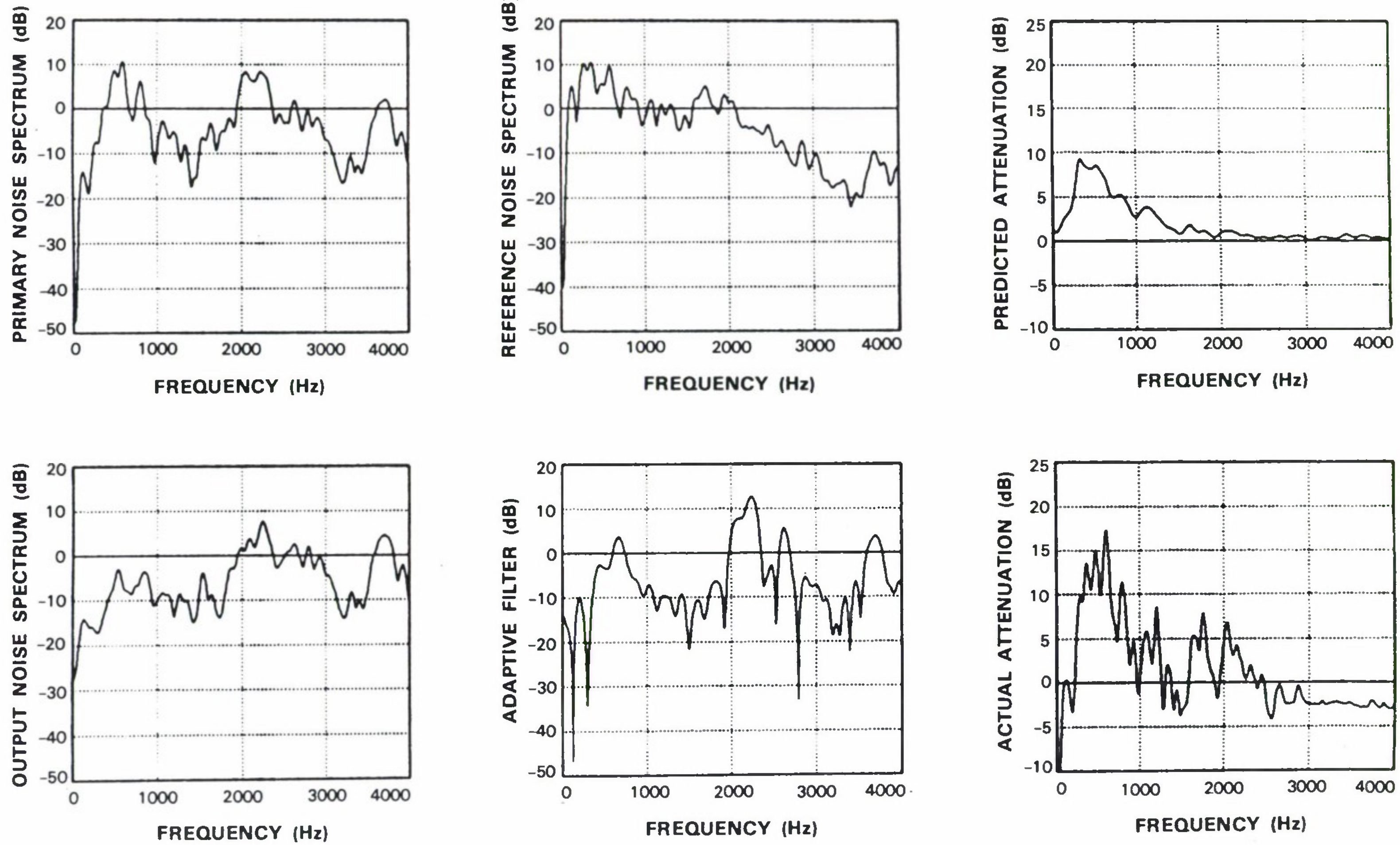


Figure 5-7. Spectral analysis: LL3.

76219-14

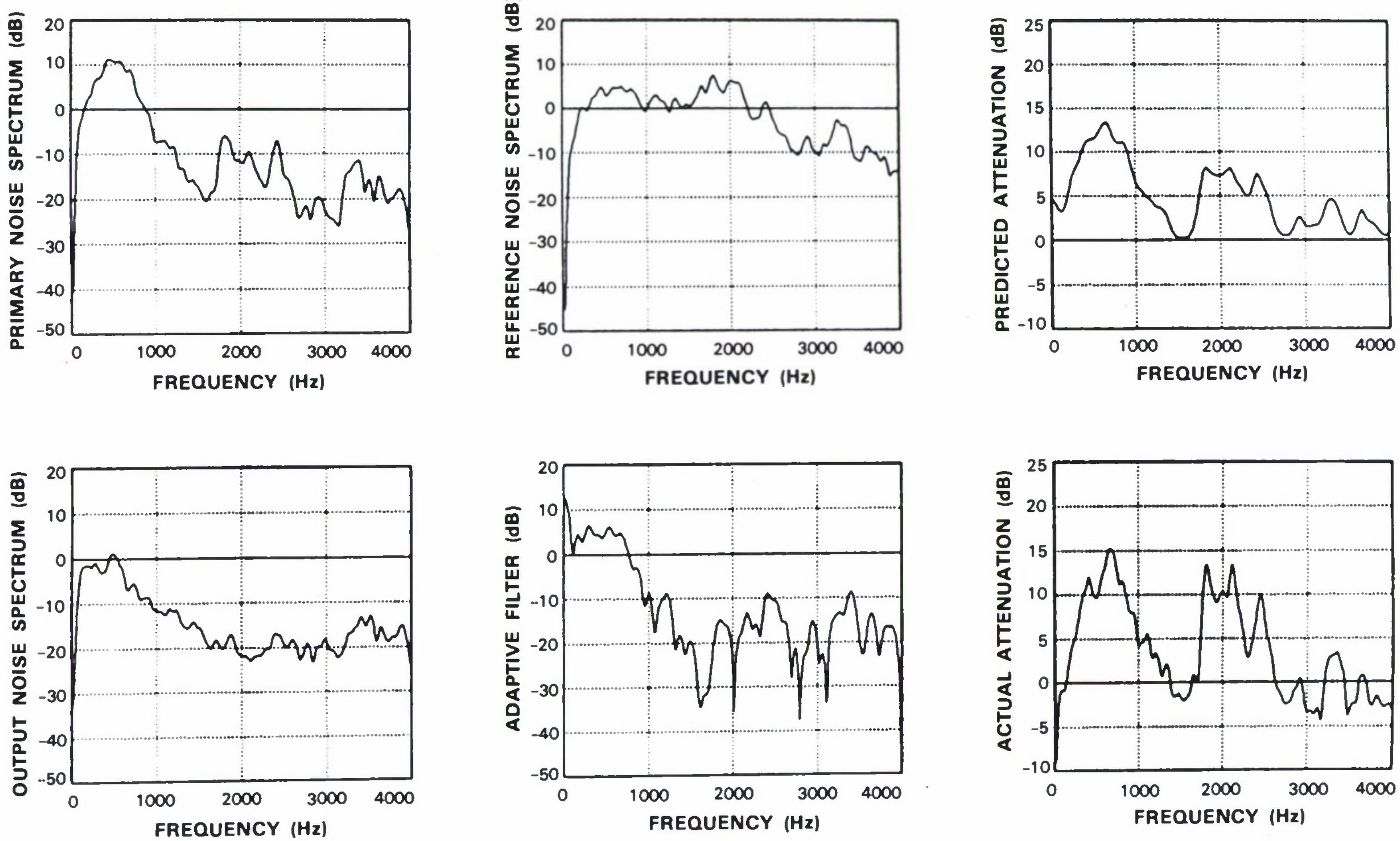


Figure 5-8. Spectral analysis: LL4.

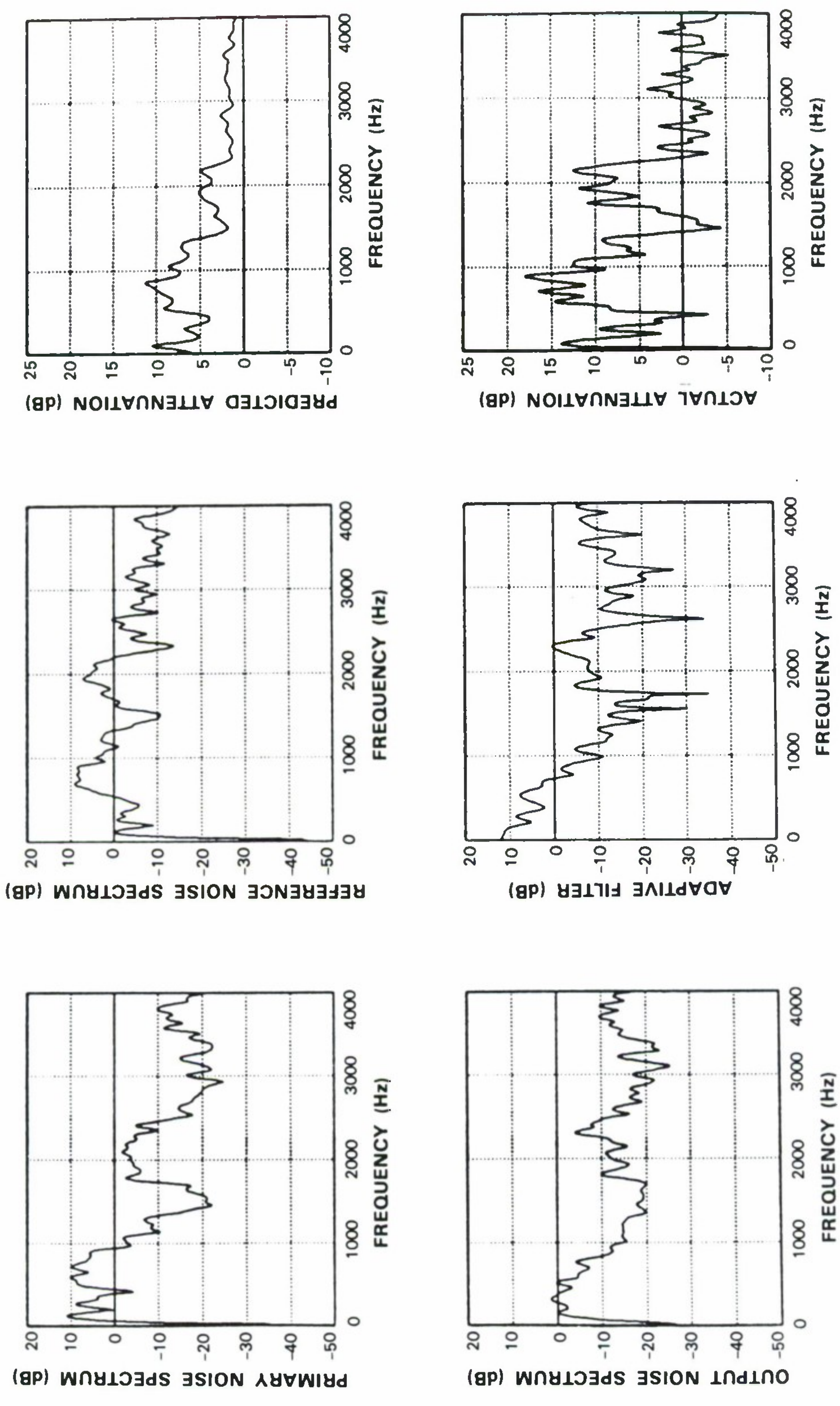


Figure 5-9. Spectral analysis: LL5.

76219-16

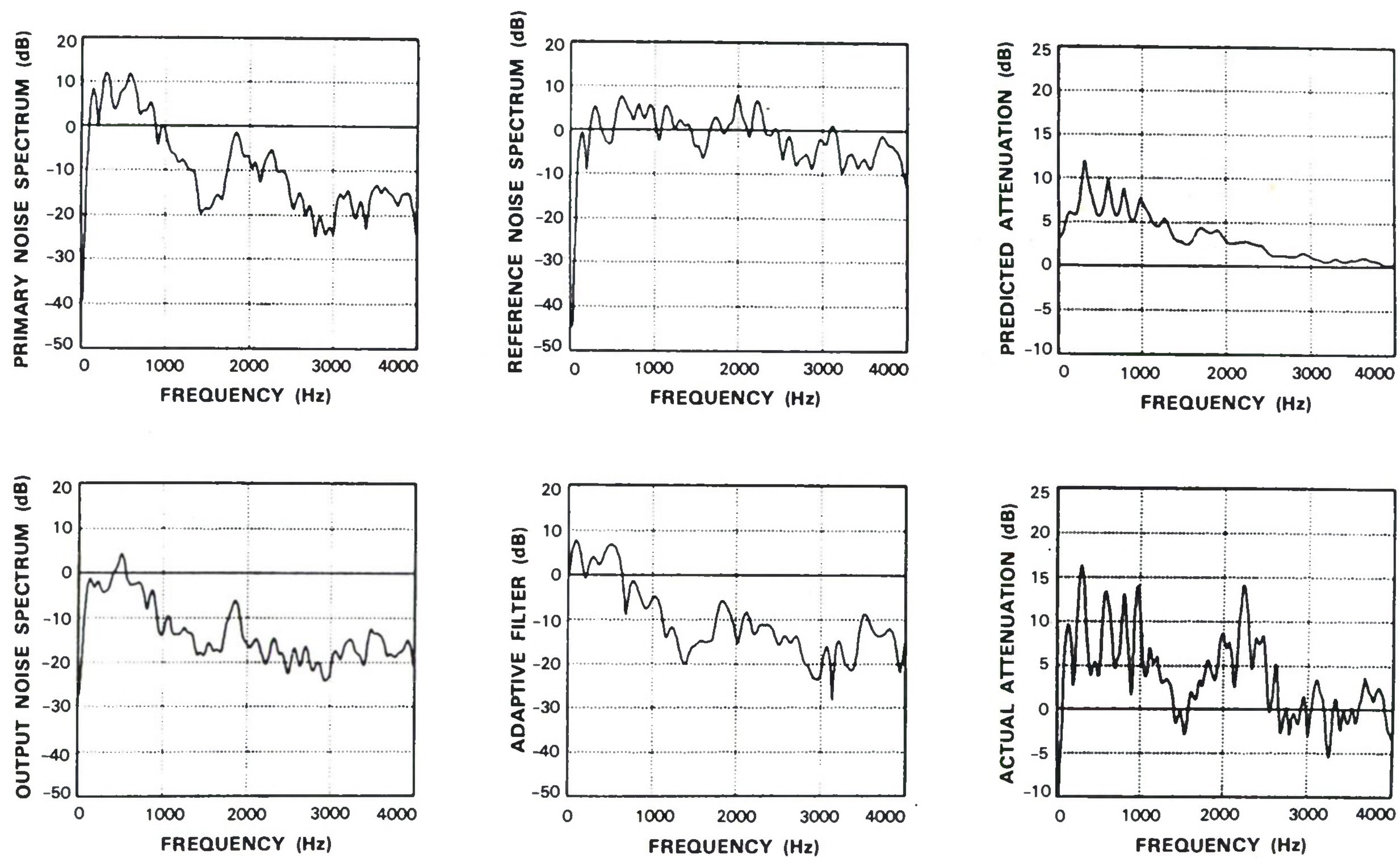


Figure 5-10. Spectral analysis: LL6.

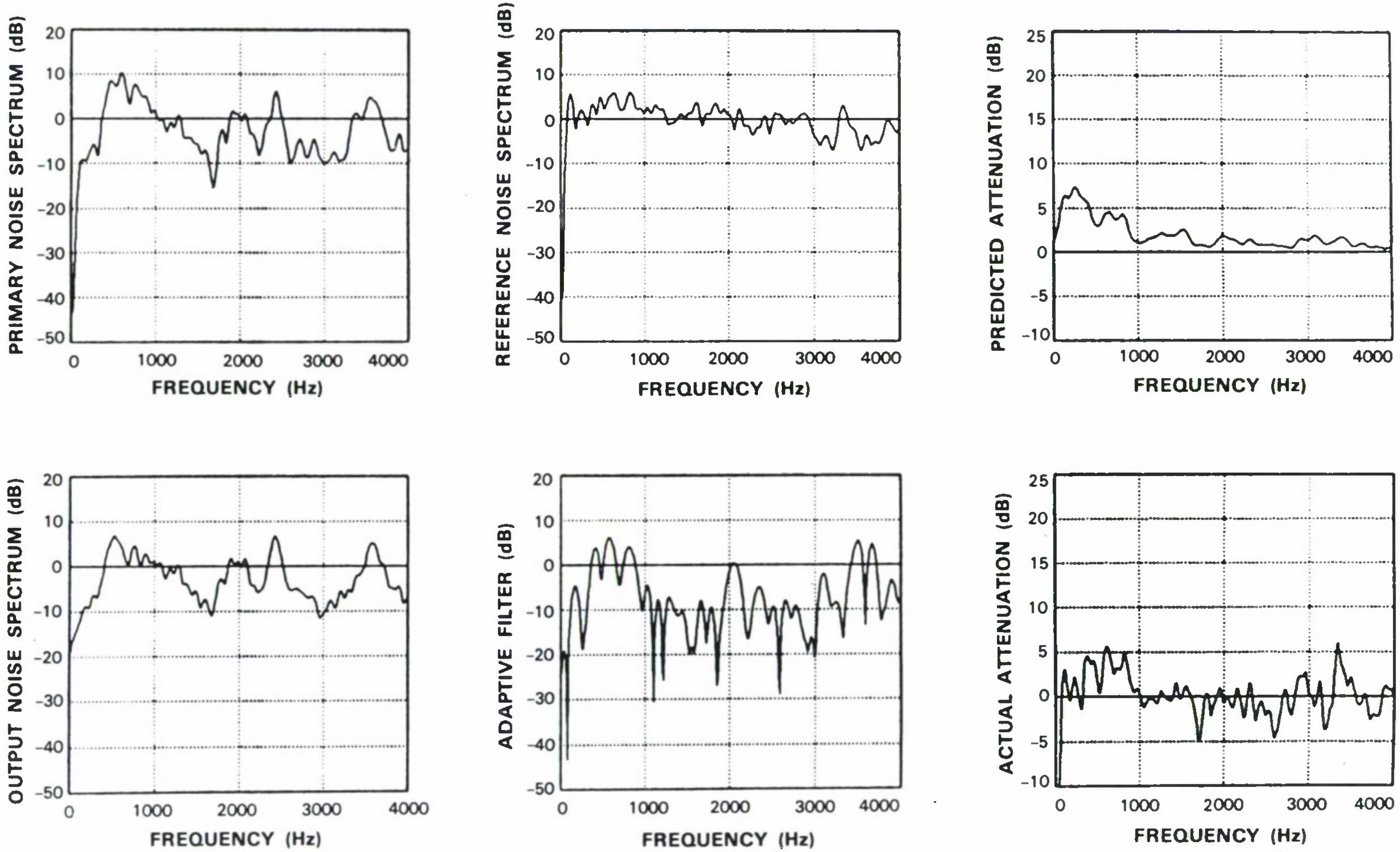


Figure 5-11. Spectral analysis: LL7.

76213-18

graph of the transfer function of the adaptive filter. (In this particular example, the adaptive filter is primarily tracking the transfer function of the face mask.) With the primary signal so heavily concentrated at low frequencies, it is not surprising that ANC performed so well (15 dB noise reduction) in this case. In the discussion of Darlington's work in Section 3.3, it was shown that, in a diffuse field, ANC should be expected to perform well only if the primary noise spectrum is concentrated below about 1 kHz. A few comments should now be made about graphs of the predicted attenuation function and the actual attenuation function. From the graph of the predicted attenuation, we see that the expected noise reduction is large at frequencies below 1 kHz, and small at higher frequencies. Equivalently, the coherence is large below 1 kHz, and smaller at high frequencies. (Because the coherence always lies between zero and one, the predicted attenuation function is always nonnegative.) The actual attenuation function exhibits the same type of behavior. However, below 1 kHz, the actual attenuation is somewhat better than predicted. Above 3 kHz, the actual attenuation is worse than predicted. Indeed, the actual attenuation is negative above 3 kHz. Therefore, the predicted attenuation function is only a rough estimate of the actual attenuation. With this greater understanding of Harrison's data, let us now look at the data from the other experiments, in hopes of explaining the variations in ANC performance.

The most striking result of this spectral analysis involves the primary noise spectra. In the cases where an omnidirectional primary microphone was used (LL4, LL5, and LL6), the primary noise spectra are concentrated mostly at low frequencies (see Figures 5-8, 5-9, and 5-10). For example, in LL4, the frequency band below 1 kHz comprises 97 percent of the primary noise energy. On the other hand, in the cases where a gradient primary microphone was used (LL1, LL2, LL3, and LL7), the primary noise spectra have a much broader shape, with large concentrations near 500 Hz and 2 kHz (see Figures 5-2, 5-3, 5-4, and 5-11). For example, in LL1, the frequency band contains only 37 percent of the primary noise energy. This is a direct result of the noise canceling behavior of the gradient microphone. As discussed in Section 3.1, the gradient microphone offers significant attenuation of the far-field noise at low frequencies. Equivalently, this can be viewed as a relative boost of the high-frequency energy. The transfer function of the adaptive filter also exhibits this high-frequency boost in the cases where a gradient primary microphone was used. (In this case, the adaptive filter tends to track the cascade of the transfer function of the face mask and the transfer function of the primary microphone.)

Therefore, it is clear that the use of a gradient microphone results in a wideband primary noise spectrum, whereas the use of an omnidirectional microphone results in a primary noise spectrum with very little energy above 1 kHz. Looking again at Figure 5-1, we see that the primary spectrum and the adaptive filter transfer function most closely resemble the ones corresponding to the experiments that used an omnidirectional primary microphone. This is strong evidence that the primary sensor in Harrison's experiment was an omnidirectional microphone.

Examination of the attenuation functions shows that, in all of the experiments, the noise reduction was greatest below 1 kHz. Above 1 kHz, there was little or no reduction of the noise. In fact, in many cases there was even a slight increase in the noise level at frequencies above 3 kHz. This was most pronounced when the primary noise level was very low at high frequencies (measurement noise was then a significant factor). As expected, the coherence estimate is an approximate, but useful, indicator of the performance of ANC. Most of the graphs of the predicted attenuation show that high correlation occurs only at low frequencies, in agreement with Darlington's results. It is interesting that this behavior appears not only in the multiple-loudspeaker experiments, but also in the single-loudspeaker experiments. Therefore, we can infer that the simulated noise field was fairly diffuse, even when a single loudspeaker was used. In most cases, the actual attenuation exceeded the predicted attenuation at low frequencies; at high frequencies, the actual attenuation was usually less than the predicted attenuation. Nevertheless, the coherence function and the predicted attenuation function are still reasonable predictors of the performance of ANC.

Further examination of the attenuation functions reveals another interesting point. The two experiments with the smallest attenuation are WP3 and LL7 (see Figure 5-4 and 5-11), in which multiple loudspeakers were used. Table 5-1 shows that these experiments also exhibited the least amount of noise reduction. As expected, the increased diffuseness of the noise field resulted in less coherence and less noise reduction. The WP3 data shows very little coherence, whereas the LL7 data shows slightly more coherence at low frequencies, in accordance with Darlington's results. These two experiments took place in rooms at different locations, so the difference is attenuation may be a result of acoustic differences between the two rooms.

Together, the performance measurements in Table 5-1 and observations of the spectral estimates provide a fairly consistent explanation of the behavior of ANC in the cockpit environment. Based on this discussion, it should now be clear that, as predicted, noise reduction is only effective at frequencies below about 1 kHz. Consequently, ANC performed best when the primary noise spectrum was concentrated at low frequencies. This low-frequency concentration was strongest in the experiments that used an omnidirectional primary microphone. Furthermore, there is convincing evidence that an omnidirectional primary microphone was used in Harrison's experiment. This perhaps explains why ANC performed so well with his data. In fact, Table 5-1 shows that ANC performed slightly better in HAR than in LL4, LL5, and LL6, the other experiments that used an omnidirectional microphone. This is explained by the differences in the reference noise spectra. In HAR, the reference noise spectrum tilts downward by about 10 dB/octave, whereas the reference noise spectra for LL4, LL5, and LL6 are more flat. Therefore, the primary noise spectrum for HAR has even less high-frequency energy than the primary spectra for LL4, LL5, and LL6. Because of this increased low-frequency concentration, the amount of noise reduction was slightly higher.

6. CONCLUSION

6.1 SUMMARY OF RESULTS

Previous research of adaptive noise cancellation in a fighter jet cockpit environment has been inconclusive. In Harrison's cockpit simulation, ANC yielded promising results. However, Darlington, et al., later claimed that Harrison's success was due to a lack of diffuseness in the simulated noise field. Because of this and other shortcomings of the simulation, there still has been much uncertainty about the effectiveness of ANC in an actual cockpit environment. A series of new experiments were performed in an effort to resolve these questions.

Altogether, eleven sets of data (primary and reference noise signals) were studied, including a copy of Harrison's experimental data. Because the main issue was the effect of the noise field on ANC performance, it was sufficient to consider only the noise signals. Therefore, the speech signals were not included in the analysis.

In the first phase of data analysis, the performance of ANC was measured for each experiment. Significant noise reduction was observed only in the experiments in which an omnidirectional primary microphone was used. The worst performance (about 1 dB attenuation) was observed in the two experiments in which multiple loudspeakers were used. This supports Darlington's prediction that the performance should decrease as the diffuseness of the noise increases. Therefore, in actual cockpit environment, with speech present and with nonideal conditions, the noise reduction would be negligible.

In the second phase of processing, a spectral analysis of the data was performed. Based on this spectral analysis, several conclusions can be drawn. First, the primary noise signal contained much more high-frequency energy when a gradient primary microphone was used than when an omnidirectional primary microphone was used. This distinction provides evidence that the primary sensor in Harrison's experiment was an omnidirectional microphone. Second, examination of the power spectra shows that, in most cases, significant noise reduction only appears below about 1 kHz. Together, these two observations explain why the performance of ANC was best when an omnidirectional primary noise spectrum was concentrated at low frequencies.

Although the use of an omnidirectional primary microphone does result in better ANC performance, this is somewhat misleading. One must keep in mind that, in this case, ANC is canceling the same noise components that a gradient microphone would cancel, namely, the band of frequencies below 1 kHz. Therefore ANC with an omnidirectional primary microphone is no better than use of a gradient microphone with no ANC. In fact, the gradient microphone attenuates the low-frequency noise even more than ANC does. The gradient microphone is preferred for other reasons as well. As discussed in Section 3.1, the gradient microphone tends to counteract the speech distortion caused by the small enclosure of the face mask. In conclusion, it appears that single-reference ANC would be ineffective in an actual cockpit environment.

6.2 RECOMMENDATIONS FOR FUTURE RESEARCH

In regard to future research, two issues deserve mentioning. First, although single-reference adaptive noise cancellation appears to be ineffective, it remains uncertain whether multiple-reference ANC would be effective. However, it is clear that poor coherence would still be a problem in the multiple-reference case. With a single reference sensor in a diffuse noise field, the primary noise signal and the reference noise signal exhibit very little coherence above about 1 kHz. Additional reference signals would suffer from the same problem. In order for multiple-reference ANC to be successful, the coherent portions (i.e., the portions coherent with the primary noise) of the reference signals would have to somehow combine constructively to form a signal more coherent with the primary noise.

However, a more fundamental question is whether further noise reduction is really that important. The premise has been that vocoders and speech recognition systems fail primarily because of excessive noise interference. However, with a gradient microphone, the primary signal-to-noise ratio is about 25 dB, which is really not so bad. Furthermore, recent research[15] indicates that noise interference inside the face mask is not the major source of degradation in speech recognition systems. The variations in the pilot's speech appear to be a much more significant problem. It has been established[17] that large noise levels at a person's ears, as well as mental and physical stress, can dramatically alter speech production. In particular, there is an increase in pitch and emphasis of high-frequency components. This phenomenon and its effect on speech recognition is an important concern that deserves further investigation.

ACKNOWLEDGMENTS

First, I would like to thank Professor Jae Lim for his suggestion of the research topic and for his helpful guidance during the course of this work.

Also, I am very grateful to Elliot Singer, at MIT Lincoln Laboratory, for his encouragement and many helpful discussions. His participation in the many tedious recording sessions deserves many thanks.

In addition, I would like to thank Bill Harrison for providing a copy of his experimental data. His cooperation was very helpful.

The primary facilities used in this research were provided by the Digital Signal Processing Group of the MIT Research Laboratory of Electronics and by Group 24 at MIT Lincoln Laboratory. It has been a pleasure working with both groups. Experimental facilities for obtaining recordings in a simulated aircraft noise environment were also provided by the Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base; the cooperation and assistance of Richard McKinley of AMRL is gratefully acknowledged.

Support for the research work and experimental facilities was provided in part by the Rome Air Development Center (RADC/EEV) under contract F19628-K-0028. The Lincoln Laboratory experimental facilities were supported by RADC/EEV under contract F19628-85-C-0002. This work was also supported by a National Science Foundation Graduate Fellowship.

REFERENCES

1. S. Boll and D. Pulsipher, IEEE Trans. Acoust. Speech Signal Process. **28**, 752 (1980).
2. G. Carter, C. Knapp, and A. Nuttall, IEEE Trans. Audio Electroacoust. **21**, 337 (1973).
3. C. Cowan and P. Grand, Eds., *Adaptive Filters* (Prentice-Hall, Englewood Cliffs, New Jersey, 1985).
4. P. Darlington, P. Wheeler, and G. Powell, Proc. IEEE Int'l. Conf. Acoust. Speech Signal Process., Vol. 2 (March 1985), pp. 716-719.
5. W. Gardner, Signal Process. **6**, 113 (1984).
6. M. Gayford, *Electroacoustics: Microphones, Earphones, and Loudspeakers* (Elsevier, New York, 1971).
7. W. Harrison, J. Lim, and E. Singer, Proc. IEEE Int'l. Conf. Acoust. Speech Signal Process., Vol. 2 (March 1984), pp. 18A.4.1-18A.4.4.
8. W. Harrison, J. Lim, and E. Singer, IEEE Trans. Acoust. Speech Signal Process. **34**, 21 (1986).
9. W. Harrison, "Speech Enhancement Using Multiple Microphones," S.M. Thesis, Dept. of Electrical Engineering, Cambridge, Massachusetts, MIT (1984).
10. M. Honig and D. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*, (Kluwer Academic Publishers, Massachusetts, 1984).
11. R. Miller *et. al.*, "Simulation of Cockpit Noise Environments in Four Tactical Aircraft for the Purpose of Testing Speech Intelligibility," Report No. 4465, Bolt Beranek and Newman (1981).
12. C. Morrow and A. Brouns, J. Acoust. Soc. Am. **50**, 1 (1971).
13. C. Morrow and A. Brouns, J. Acoust. Soc. Am. **50**, 10 (1971).
14. A. Nuttall and G. Carter, IEEE Trans. Acoust. Speech Signal Process. **24**, 582 (1976).
15. D. Paul *et al.*, "Robust HMM-Based Techniques for Recognition of Speech Produced Under Stress and in Noise," in *Proceedings of Speech Technology '86* (Media Dimensions, New York, April 1986), pp. 241-249.
16. A Piersol, J. Sound Vib. **56**, 215 (1978).

17. D. Pisoni *et al.*, Proc. IEEE Int'l. Conf. Acoust. Speech Signal Proc., Vol. 4 (March 1985), pp. 1581-1584.

18. P. Welch, IEEE Trans. Audio Electroacoust. **15**, 70 (1967).

19. B. Widrow *et al.*, Proc. IEEE **63**, 1692 (1975).

20. B. Widrow, S. Sterns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, New Jersey, 1985).

21. B. Widrow *et al.*, Proc. IEEE **64**, 1151 (1976).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ESD-TR-86-096	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Adaptive Noise Reduction in Aircraft Communication Systems		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER Technical Report 756
7. AUTHOR*(s)* Jeffrey J. Rodriguez		8. CONTRACT OR GRANT NUMBER*(s)* F19628-85-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, MIT P.O. Box 73 Lexington, MA 02173-0073		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element Nos. 33401F and 64754F
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Systems Command, USAF Andrews AFB Washington, DC 20334		12. REPORT DATE 20 January 1987
		13. NUMBER OF PAGES 60
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Electronic Systems Division Hanscom AFB, MA 01731		15. SECURITY CLASS. *(of this Report)* Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

Approved for public release; distribution unlimited.

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

adaptive noise cancellation
speech enhancement
adaptive filters
noise reduction
digital speech processing
noise-canceling microphones
speech processing

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

In many military environments, such as fighter jet cockpits, the increasing use of digital communication systems has created a need for robust vocoders and speech recognition systems. However, the high level of ambient noise in such environments makes vocoders less intelligible and makes reliable speech recognition more difficult. One method of enhancing the noise-corrupted speech is adaptive noise cancellation. In previous research, this method was tested in a simulated cockpit environment, yielding impressive results. However, in new simulations, reflecting more realistic conditions, adaptive noise cancellation has been less successful. Spectral analysis of the data shows that the spectral concentration of the ambient noise, along with the microphone characteristics, has a significant effect on the performance of adaptive noise cancellation.